A.I.&D.S. Technical Report TR-1048-01

Multitarget Multisensor Tracking Problems - Part I: A General Solution and a Unified View on Bayesian Approaches

Shozo Mori, Chee-Yee Chong, Edison Tse and Richard P. Wishner

June 1983
August 1984 (Revised)

ADVANCED INFORMATION & DECISION SYSTEMS
Mountain View, CA 94040

AI&DS

A.I.&D.S. Technical Report TR-1048-01

Multitarget Multisensor Tracking Problems - Part I: A General Solution and a Unified View on Bayesian Approaches

Shozo Mori, Chee-Yee Chong, Edison Tse and Richard P. Wishner

June 1983
August 1984 (Revised)

ADVANCED INFORMATION & DECISION SYSTEMS
Mountain View, CA 94040

PREFACE

This report was originally written as a manuscript for a paper submitted to *IEEE Transaction on Automatic Control* in June 1983.* In the process of modifying the manuscript, we had to condense its technical discussions considerably and omit many details. One of our motivations for writing the paper, however, was our recognition that the multitarget tracking problem is a non-conventional estimation problem and needs a new mathematically sound foundation. Thus we intended in this report to present the complete mathematical descriptions which we could not include in the article. Naturally, our main objective in this report was to make every step in theoretical development as clear and rigorous as possible, even at the expense of some readability. The two major results are stated in the form of two theorems. Several details have been added in order to ensure theoretical completeness. The review of the recently published NRL report by Dr. I.R. Goodman has been added. In effect, this report summarizes our efforts at A.I.&D.S.**, from 1980 to 1984, to create a general theory of multitarget multisensor tracking. Our original intention of writing a two-part paper remains; in the near future we will complete Part II, which will contain implementational issues such as hypothesis management with several illustrative examples.

The theoretical development for this report was mainly supported by the Defense Advanced Research Project Agency under contract MDA903-81-0333. The documentation preparation was supported by the continued contract MDA903-83-C-0333.

August 1984, Shozo Mori

* S. Mori, C.-Y. Chong, E. Tse and R. P. Wishner, "Tracking and Classifying Multiple Targets Without *A Priori* Identification." *IEEE Trans. on Auto. Contr.*, Vol. AC-31, No. 5, May 1986.

** In 1985 Advanced Information & Decision Systems changed its name to Advanced Decision Systems (ADS).

Second Preface

A final draft of this report was completed in August, 1984, as mentioned above. Due to very unusual circumstance, it has remained in that state for almost three years. Since we have cited this report in other papers, we are obliged to publish it. I hope that this process will be finalized in teh immediate future.

April 1987, Shozo Mori

ACKNOWLEDGEMENT

The authors wish to thank Mr. Kuo-Chu Chang for many valuable comments.

ABSTRACT

Based upon a general target-sensor model which allows dependence among targets and state-dependent target detection, a Bayesian solution to the multitarget tracking problem is derived. When this solution is applied to a special class of models, a less general but more implementationally feasible class of algorithms is obtained. Representative existing algorithms are then compared with our results. By doing so, we provide a unified view on Bayesian approaches to the multitarget tracking problem. Part I covers most of the analytical results, while in Part II, hypothesis management and other issues pertaining to implementation of multitarget algorithms are discussed with several examples.

NOMENCLATURE

Symbol	*Explanation*
$\{x \mid C\}$	the set of x satisfying condition C.
$\{x\}$	a singleton: the set whose only element is x.
$\{x_1, \cdots, x_n\}$	the set consisting of $x_1, \ldots, x_n$: $\{x_1, \cdots, x_n\} = \{x_i \mid 1 \le i \le n\}$ (When $n=0$, the set is empty.).
$\in, \cup, \cap, \subseteq, \subset$	the standard set theoretic notations: member of, union, intersection, subset of and proper subset of.
$\setminus$	the set substruction operation: $A \setminus B = \{a \in A \mid b \notin B\}$.
$\times$	the direct product: $A \times B = \{(a,b) \mid a \in A \text{ and } b \in B\}$.
$\#(\cdot)$	the cardinality of a set: $\#(E)$ is the cardinality of set E, i.e., the number of elements belonging to E.
$f: A \to B$	a function defined on set A taking values in set B.
$(a_i)_{i \in I}$, $(a_i)_{i=1}^n$	indexed tuple: theoretically equivalent to a function a defined on set I or $\{1, \ldots, n\}$.
$n!$	factorial: $n! = 1 \cdot 2 \cdots n$.
$\circ$	function composition operation: $(f \circ g)(x) = f(g(x))$.
$\chi(\cdot\,; E)$	the indicator function of set E: $\chi(e\,; E)$ is 1 if $e \in E$ 0 otherwise.
$\|x\|_A$	the semi-norm on a Euclidean space defined by a nonnegative definite symmetric matrix A: $\|x\|_A^2 = x^T A x$ where x^T is the transpose of vector x.
A_k	the assignment function at data set k: a random function defined on a subset of the target index set $I_T \triangleq \{1, \ldots, N_T\}$ taking values in $J_M(k) \triangleq \{1, \ldots, N_M(k)\}$ where N_T is the total number of targets and $N_M(k)$ is the number of measurements in data set k.
$\mathrm{A}(I\,J)$	the set of all the one-to-one functions defined on set I taking values in J.

$\mathbf{A}$	a 0-1 matrix representing the constraint in the 0-1 integer programming form of hypothesis evaluation.
$B_H^{(K)}$	the normalizing constant for the batch-processing type hypothesis evaluation for the hypotheses on the cumulative data set $Z^{(K)}$ in i.i.d.-Poisson cases.
$\tilde{B}_H^{(K)}$	the normalizing constant for the modified batch-processing type hypothesis evaluation for the hypotheses on the cumulative data set $Z^{(K)}$ in i.i.d.-Poisson cases.
$\beta_{NT}(k)$	the density of new targets (Reid's terminology in [4]).
$\beta_{FA}(k)$	the false alarm density: $\beta_{FA}(k)=\nu_{FA}(k)/\mu_{s_k}(Y_{s_k})$ where $\nu_{FA}(k)$ is the expected number of false alarms in data set k and $\mu_{s_k}(Y_{s_k})$ is the scan (search) volume of sensor s_k.
$C_H^{(k)}$	the normalizing constant for the recursive hypothesis evaluation equation at data set k in general cases.
$C_{N_T}^{(k)}$	the normalizing constant for updating the number-of-target distribution at data set k in general cases.
$C_X^{(k)}$	the normalizing constant for updating the target state distribution at data set k in general cases.
$c_k(y \mid Z_\tau)$	the normalizing constant for updating the track state distribution conditioned by a cumulative data set Z restricted by a track τ with a measurement value y in i.i.d.-Poisson cases.
$\mathbf{c}$	the objective function coefficient vector in the 0-1 integer programming form hypothesis evaluation.
$Dom(f)$	the domain of function f.
$E(\cdot), E(\cdot \mid \cdot)$	the mathematical expectation; unconditional and conditional.
$F_{\Delta\tau}^{n}(\cdot \mid \cdot)$	the transition probability of the target states in general cases when the number N_T of targets is n.
$f_{\Delta\tau}(\cdot \mid \cdot)$	the state transition probability for each individual target in i.i.d.-Poisson cases.

$g(y \mid x,k)$	the density of extended state-to-measurement transition probability at data set k: $g(y \mid x,k)$ is $p_M(y \mid x,k)p_D(x \mid k)$ if $y \neq \theta$ (real data -- target detected) $1-p_D(x \mid k)$ otherwise (no data – target undetected).
$\mathbf{H}(Z^{(K)})$	the set all the (data-to-data) hypotheses on cumulative data set $Z^{(K)}$.
H_k	the state-to-measurement matrix in linear-gaussian measurement models.
$h_k(\lambda)$	the objective function in the 0-1 linear programming form of hypothesis evaluation.
$I_{DT}(k)$	the set of all the indices for the targets detected in data set k: $I_{DT}(k)=Dom(A_k)\subseteq I_T$.
I_T	the target index set: $I_T=\{1,\ldots,N_T\}$ where N_T is the total number of targets.
J , $\bar{J}$, $J^{(k)}$	the cumulative measurement index set up to data set k (when the superscripts are dropped, J is the measurement index set for the currently available cumulative data set and $\bar{J}$ for the same minus the most recent data set).
$J_{FA}(k)$	the set of indices for all the false alarms in data set k: $J_{FA}(k)=J_M(k)\backslash Im(A_k)$
$J_M(k)$	the measurement index set at data set k: $J_M(k)=\{1,\ldots,N_M(k)\}$ where $N_M(k)$ is the number of the measurements in data set k.
$j_{FA}(m,\lambda \mid k)$	the set of indices for the false alarms in data set k hypothesized by a hypothesis λ given $N_M(k)=m$: $j_{FA}(m,\lambda \mid k)=\{j \in \{1,\ldots,m\} \mid (j,k) \notin \bigcup_{\tau \in \lambda}(\tau \mid k)\}$.
k , k', K	indices for data sets.
$\mathbf{L}_k(z(k),\lambda \mid Z^{(k-1)},\bar{\lambda})$	the likelihood of $(z(k),\lambda)$, i.e., the data set $z(k)$ and the data-to-data hypothesis λ, given the cumulative data sets $Z^{(k-1)}$ and the predecessor hypothesis $\bar{\lambda}$, in general cases.
$L_k(y \mid \bar{Z}_{\bar{T}})$	the likelihood of measurement y given a cumulative data set restricted by a track, $\bar{Z}_{\bar{T}}$, in i.i.d.-Poisson cases.

$l_K(\tau, Z^{(K)})$	the track likelihood of track τ given the cumulative data set $Z^{(K)}$ at data set K in i.i.d.-Poisson cases.
$\tilde{l}(\tau, Z^{(K)})$	the modified track likelihood of track τ given the cumulative data set $Z^{(K)}$ in i.i.d.-Poisson cases with the Poisson number-of-false-alarm assumption.
Λ_k	the random collection of sets of measurement indices in the cumulative measurement index set $J^{(k)}$, representing the measurements from the (real) targets. (Each possible realization of Λ_k is called a data-to-data hypothesis on $Z^{(k)}$. The subscript k is dropped when there is no fear of confusion.
λ , $\bar{\lambda}$	data-to-data hypotheses: $\bar{\lambda}$ is the unique immediate predecessor of λ.
$\lambda \mid Z$	the data-to-data hypothesis λ restricted by the cumulative data set Z.
λ_{new}	the set of newly initiated tracks (hypothesizing newly detected targets) in data-to-data hypothesis λ.
λ_{old}	the set of previously initiated tracks (hypothesizing previously detected targets) in data-to-data hypothesis λ.
μ_s	the hybrid measure on the measurement value space Y_s for sensor s.
$N_{DT}(k)$	the number of targets detected at data set k: $N_{DT}(k)=\#(I_{DT}(k))=\#(Dom(A_k))$.
$N_{FA}(k)$	the number of false alarms in data set k.
$N_M(k)$	the number of measurements in data set k.
$N_{NT}(k)$	the number of targets which are detected at data set k for the first time.
N_T	the total number of targets
$\bar{N}_T$	the *a priori* upper bound on the total number of targets (possibly $+\infty$ such as in i.i.d.-Poisson cases).
ν_0	the *a priori* expected number of targets in i.i.d.-Poisson cases.
$\nu_{FA}(k)$	the expected number of false alarms at data set k.

ν_k	the *a posteriori* expected number of undetected targets at data set k.
Ω_k	the function defined on Λ_k taking values in the target index set $I_T \triangleq \{1,...,N_T\}$: Ω_k uniquely assigns a target index to each τ in Λ_k. (We call every possible realization of Ω_k a target-to-track hypothesis. The subscript is dropped when there is no fear of confusion.)
ω , $\bar{\omega}$	target-to-track hypotheses: hypotheses on the realization of Ω_k: $\bar{\omega}$ is a restriction of ω.
P_D	the detection probability function in general cases.
$P_{\mathbf{H}}^{(k)}$	the *a posteriori* probability of each data-to-data hypothesis at data set k (hypothesis evaluation function).
P_M	the measurement value probability density function in general cases.
$P_{N_{FA}}$	the number-of-false alarm function in general cases.
$P_{N_T}^{(k)}$	the *a posteriori* distribution of the number N_T of targets given a data-to-data hypothesis at data set k in general cases.
$P_{\mathbf{X}}^{(k)}$	the *a posteriori* target state distribution given the number of targets and a data-to-data hypothesis at data set k in general cases.
Prob.	the probability measure on the underlying probability space.
p_D	the detection probability function in i.i.d.-Poisson cases.
$p_D^{\max}$	the maximum value of the detection probability function in i.i.d.-Poisson cases.
p_{FA}	the density of the distribution of the false-alarm values in i.i.d.-Poisson cases.
p_M	the density of the target-state-to-measurement transition probability in i.i.d.-Poisson cases.
$p_{N_{FA}}$	the number-of-false-alarm probability distribution in i.i.d.-Poisson cases.
$p_t(\cdot \mid \bar{Z}_\tau)$	the track state distribution conditioned by the cumulative data set restricted by a track, i.e., $\bar{Z}_\tau$.

Π	the n-target permutation homeomorphism.
π	a permutation on $\{1,.....,n\}$.
$Q^{(K)}$	a labeled partition (Goodman's notation in [5]).
Q_0^n	the *a priori* target state distribution when the number N_T of targets is n in general cases.
q_0	the *a priori* target state distribution common to all the targets in i.i.d.-Poisson cases.
R_k	the error variance matrix in i.i.d.-Poisson linear-gaussian cases.
$\mathbb{R}$	the set of real numbers.
S	the set of sensors.
$\bar{s}$	a super-sensor (used to explain the PDA and JPDA algorithms).
s_k	the sensor which generates the k-th data set.
$\Sigma_{k\|k-1}$	the target state prediction variance in i.i.d.-Poisson-gaussian cases.
$\mathbf{T}(Z^{(k)})$	the set of all the tracks on the cumulative data set $Z^{(k)}$.
t_k	the time at which the k-th data set is generated.
θ	the symbol for "nothing," "no target present," or "no measurement generated.": according to the usual convention, the set of all the functions defined on the empty set taking values in a nonempty set consists of a single element whose domain and range are both empty. θ is a generic symbol for such a special function.
τ	a track: a subset of a cumulative measurement index set such that it contains at most one measurement index set for each data set.
$\mathbf{W}(\lambda,n)$	the set of all the target-to-track hypotheses consistent with a data-to-data hypothesis λ and the condition that N_T of targets is n.
$\mathbf{X}$	the common space for the individual components of target states in i.i.d.-Poisson cases.
$\mathbf{X}_n$	the target state component when the number N_T of targets is n in general cases.

$\mathbf{X}_n^C$ — the space for the state component common to all the targets when the number N_T of targets is n in general cases.

$\mathbf{X}_n^I$ — the space for the collection of individual components of target states when the number N_T of targets is n in general cases.

$\mathbf{X}_n^i$ — the individual target state space: $\mathbf{X}_n^I = \mathbf{X}_n^i \times \cdots \times \mathbf{X}_n^i$ (n *times*), in general cases.

$X(t)$ — the target states at time t in general cases.

$x_i(t)$ — the state of the i-th target in i.i.d.-Poisson cases.

$\hat{x}_{k|k-1}$ — the predicted target state in the i.i.d.-Poisson-gaussian cases.

$\mathbf{x}$ — the 0-1 vector representation of a data-to-data hypothesis in the 0-1 integer programming form of batch-processing hypothesis evaluation.

Y_s — the measurement value space for sensor s.

$y[\tau|k]$ — the measurement value assigned to track τ at data set k: $y[\tau|k]$ is $y_j(k)$ (the value of the j-th measurement) if $(j,k) \in \tau$ and is θ otherwise (there is no j such that $(j,k) \in \tau$).

$y_j(k)$ — the measurement value of the j-th measurement in the k-th data set.

Z, $\bar{Z}$, $Z^{(k)}$ — cumulative data sets: $Z^{(k)}$ is the cumulative data set available at k. When the super scripts are omitted, Z represents a currently available cumulative data set and $\bar{Z}$ is the same minus the most current data set.

Z_τ — the cumulative data set restricted by track τ.
$Z_\tau = \{(y[\tau|k],k) \mid (z(k),k) \in Z\}$.

$z(k)$ — the k-th data set: $z(k) = ((y_j(k))_{j=1}^{N_M(k)}, N_M(k), t_k, s_k)$, where $N_M(k)$ is the number of measurements and $y_j(k)$ is the value of the j-th measurement in the data set. The k-th data set is generated by sensor s_k at time t_k. The k-th data set is occasionally simply called data set k.

$\mathbb{Z}_+$ — the set of all the nonnegative integers.

TABLE OF CONTENTS

1. INTRODUCTION

Multitarget tracking is a problem for tracking and classifying a (generally unknown) number of homogeneous and/or heterogeneous objects (called targets as a generic name), based on origin-ambiguous noisy measurements from homogeneous and/or heterogeneous sensors, where some targets may not be detected and some measurements may not originate from any object of interest (false alarms, clutters, etc.). When the origin of each measurement is known, tracking and classifying are reduced to joint estimation of continuous and discrete (respectively) parts of states of stochastic processes based on discrete-time observation. Multitarget tracking problems are technically very interesting because they may require formalism which cannot be found in traditional estimation problems. They have great importance due to their wide range of applications such as air defense, air traffic control, ocean/battlefield surveillance, etc. In the past two decades, this new field has attracted numerous researchers. Past achievements are well documented in the survey paper [1] and the Naval Ocean Surveillance Correlation Handbooks, [2] and [3]. The introductory section of [4] contains a short but excellent survey and a recently published report [5] discusses representative past results extensively. Despite considerable efforts spent in this area, the present stage may well be characterized as an unorganized collections of numerous "named" or "unnamed" algorithms. As correctly pointed out in [5], no general multitarget tracking concept has been created. This failure may be attributed to the general trend that many algorithms are developed in very problem-specific ways and based on special assumptions. One of the objectives of this report is to create a set of basic concepts upon which researchers can communicate with each other.

In a recent paper [6], an attempt was made to create a unified view on this subject by describing the multitarget tracking problems in terms of a special class of dynamical systems, i.e., event-driven linear stochastic systems. In multitarget tracking problems, however, when target detection and measurement assignment to targets are considered as an "event" process, such a process is clearly driven by the target state process, not vice versa. In this report, we will present a completely different view on the subject. We contend that the multitarget tracking problem is "unconventional" or "non-classical" in the sense that it calls for a new type of mathematical modeling and formalism. The essential features of the problem are: (1) The number of the objects to be estimated is, in general, random and unknown. The number of measurements in each sensor output is random and a part of observation information. (2) Generally, there is no *a priori* labeling[1] of targets and the order of measurements in any sensor output does not contain any useful information. For example, a measurement couple (y_1,y_2) from a sensor is totally equivalent to (y_2,y_1). When a target is detected for the first time and we know it is one of n targets which have never been seen before, the probability that the measurement originating from a particular target is the same for any such target, i.e., it is $1/n$. The above properties (1) and (2) are properly reflected when both targets and sensor measurements are considered as *random sets* as defined in [7]. We will tentatively call such properties *random-set nature*. In short, according to our view, one of the fundamental aspects of multitarget tracking

[1] By this, we mean that the targets does not have *a priori* identification. See the last remark in Section 2.

problems is random-set nature. The uncertainty of the origin of each measurement in every sensor output should then be imbedded in a sensor model as a stochastic mechanism which converts a random set (set of targets) into another random set (sensor outputs).

Theories of random sets or stochastic geometry are mainly concerned with random sets whose realizations[2] are uncountable sets such as closed, open or convex sets in Euclidean spaces, and are mathematically highly sophisticated. Fortunately, when we restrict ourselves to random sets whose realizations have only finite members with probability one, we can still apply standard probabilistic techniques. For example, a random finite set X of reals can be probabilistically completely described by specifying probability $Prob.\{\#(X)=n\}$[3] for each nonnegative n and joint probability distribution with density $p_n(x_1,\ldots,x_n)$ of elements of the set for each positive n. In order for this specification to be appropriate, however, we must require every p_n to be interchangeable (permutable). This is the basic approach which we take in this report. As in almost all the existing literature on multitarget tracking, the basic task is to hypothesize the origin of each measurement and to evaluate all the possible hypotheses, or in other words, to generate and evaluate all the possible partitions of measurements. To accomplish this, we will define hypotheses and tracks (two terms widely used but often vaguely defined) rigorously and relate them to particular random sets.

In the next section, we will describe a general class of target/sensor models which form the basis for the new general multitracking algorithm described in Section 4. Section 3 gives definitions of hypotheses and tracks. For a problem statement, Sections 2 and 3 are quite lengthy, but necessary so, since to date there has been no mathematically rigorous problem statement for general multitarget tracking problems, and some confusion seems to exist in basic modeling. In Section 4, a general recursive formula for evaluating hypotheses is derived based on a general model. In Section 5, we restrict ourselves to a more limited class of models, i.e., what we may call i.i.d. (independent, identically distributed) - Poisson (number of targets) cases. The results derived in Section 5 are, however, general enough to include many existing algorithms roughly speaking as a subset, as seen in Section 6 where we compare our results with other existing algorithms. Sections 5 and 6 together provide a set of implementationally feasible algorithms and a unified view on existing algorithms. Part I covers most of the theoretical issues, whereas Part II of this report describes hypothesis management and other implementational issues with simple examples.

---

[2] In this report, a *realization* of a random element is synonymous with a *sample*, i.e., a particular value of a random element when a point in the underlying probability space is specified.

[3] For any set A, $\#(A)$ is the cardinality of A or the number of members of A.

2. MODELS OF TARGETS AND SENSORS

2.1. Target Model A *target* is a generic name for the least unit of object which we wish to track and/or classify and which, when detected by a sensor at a certain time, may generate a measurement in the sensor's output. In our model, all the targets of interest are modeled as one entity rather than as a collection of individual targets. Formally, a *target system state* at time t is a realization $(X(t),N_T(t))$ at t of a continuous-time stochastic process[4] $(X(t),N_T(t))_{t\in[t_0,\infty)}$ on a *target system state space* $\bigcup_{n=0}^{\infty}\left(\mathbf{X}_n\times\{n\}\right)$ which is the formal disjoint union of a system $(\mathbf{X}_n)_{n=1}^{\infty}$ of hybrid sets. By a hybrid set, in this report, we mean the direct product of a measurable set (called continuous part) in a Euclidean space and a finite set (called discrete part). The use of hybrid sets allows us to include discrete component into the target states, e.g., different types of targets (with different dynamics), sudden structural changes in dynamics (such as maneuvers), changes in operational modes (such as stop/go modes), etc., in addition to the usual physical (geolocational) states such as position, altitude, velocity and acceleration. The second element, $N_T(t)$, represents the total number of targets in the system at time t. When $n=0$, $\mathbf{X}_n$ is defined as $\{\theta\}$ where θ is merely a symbol for "no target" and $\theta\notin\mathbf{X}_n$ for all $n>0$. The following assumption is made for the stochastic process $(X(t),N_T(t))_{t\in[t_0,\infty)}$:

Assumption 1: (Constant Number of Targets and Markovian Property) The second component of the stochastic process (X,N_T), i.e., the number of targets N_T, is constant but in general random with a known probability distribution. For each positive integer n, given $N_T=n$, the first component X of (X,N_T) is a temporally homogeneous Markov process on $\mathbf{X}_n$ having initial distribution

$$Q_0^n(dX)=Prob.\{X(t_0)\in dX\mid N_T=n\} \tag{1}$$

and transition probability

$$F_{\Delta t}^n(dX\mid X)=Prob.\{X(t+\Delta t)\in dX\mid X(t)=X,N_T=n\} \tag{2}$$

for each $X\in\mathbf{X}_n$, each $t\in[t_0,\infty)$ and each $\Delta t\geq 0$.

Remark 1: The temporally homogeneity assumption can easily be removed. However, this assumption reduces notational complexity in the following discussions. One should not confuse the constant number, N_T, of targets with the number of detected targets or that of targets in

[4] t_0 is the time when no sensor has begun operating yet.

the sensors' fields of view. For example, when the targets are born and die as in the models described in [5] and [8], we can imbed an appropriate birth-death process into the target dynamics. In such a case, N_T is the total number of all the targets which once exists in a nonempty interval included in $[t_0,\infty)$. Implicitly, we have assumed the finiteness, i.e., *Prob.* $(N_T<\infty)=1$, which correctly reflects the reality[5], while an upper bound $\bar{N}_T$ such that *Prob.* $\{N_T\leq\bar{N}_T\}=1$ may or may not exist.

In order to model the random-set nature of the targets as a whole, we need some additional structure in the model: First, for each positive integer n, we assume that the component $\mathbf{X}_n$ of the target system space is further decomposed as

$$\mathbf{X}_n = \mathbf{X}_n^C \times \mathbf{X}_n^I \tag{3}$$

where $\mathbf{X}_n^C$ is the space for the *common target state* for all the targets in the system and $\mathbf{X}_n^I=\mathbf{X}_1^i\times\cdots\times\mathbf{X}_n^i$ (*n times*) is the direct product of n identical *individual target state spaces* $\mathbf{X}_n^i$. In a simple example, $\mathbf{X}_n^C=\{\theta\}$ (no common target state space) and every $\mathbf{X}_n^i$ is a hybrid space so that, for each n and t, when $N_T=n$, $X(t)=(x_1(t),\ldots,x_n(t))$ and each $x_i(t)$ represents the i-th target's state in $\mathbf{X}_n^i$ at time t. In another example, $\mathbf{X}_n^C=\mathbf{X}_n^i=\mathbb{R}^2$ (= the set of pairs of reals)[6], and $(x_0,x_1,\ldots,x_n)$ in $\mathbf{X}_n$ represents a target system state for a group of targets, where x_0 is the hypothetical centroid and each x_i is the displacement of target i from x_0. We call a function $\Pi:\mathbf{X}_n\rightarrow\mathbf{X}_n$ an *n-target permutation homeomorphism (induced by a permutation π)* if, for every $(X_C,(X_i)_{i=1}^n)$ in $\mathbf{X}_n=\mathbf{X}_n^C\times\mathbf{X}_n^I$, we have

$$\Pi(X_C,(X_i)_{i=1}^n)=(X_C,(X_{\pi(i)})_{i=1}^n) \tag{4}$$

for a permutation π on $\{1,\ldots,n\}$. Then we can state our interchangeability requirement for targets in the following way:

Assumption 2: (Target Interchangeability (1)) For each positive integer n, initial distribution Q_0^n and state transition probability $F_{\Delta t}^n$ are *interchangeable* (or *permutable*) with respect to the individual target state part $\mathbf{X}_n^I$ of $\mathbf{X}_n$, i.e.,

$$Q_0^n(\Pi(dX))=Q_0^n(dX) \tag{5}$$

---

[5] Consider all the airplanes in the world. The number of the airplanes which have been built by the present time is of course finite. The number of airplanes which will be built in the future is also finite although a probability-one upper bound on it may not exist.

[6] $\mathbb{R}=(-\infty,\infty)$

and

$$F^n_{\Delta t}(\Pi(dX)\,|\,\Pi(X)) = F^n_{\Delta t}(dX\,|\,X)\ , \tag{6}$$

for any n-target permutation homeomorphism Π.

This assumption states that there is no *a priori* labeling[7] of targets and the labeling by means of integers, $\{1,..,n\}$, is just one of many equivalent ways of labeling. Since targets do not have *a priori* labels, sensors can not treat targets with *a priori* discrimination. This fact must be incorporated into any sensor model as will be done below. Thus our target model is represented by $(Prob.\{N_T=n\}, Q^n_0, (F^n_{\Delta t})_{\Delta t \geq 0})^{\infty}_{n=0}$ which satisfies *Assumptions 1 and 2.*

2.2. Sensor Model Let S be a finite set of sensors. Each sensor s in S is modeled as a generic mechanism which observes the target system state space and generates a finite set of *measurements*, called *data sets*[8], intermittently according to a certain sampling pattern. Each measurement in a data set output from sensor s in S is an element of *measurement value space* Y_s for sensor s. Each Y_s is also a hybrid space with hybrid measure μ_s. By the hybrid measure on a hybrid space, we mean the direct product measure of Lebesgue measure on the continuous part and the counting measure on the discrete part. The continuous part of Y_s may be used for analog information such as range, azimuth, elevation, etc., whereas the discrete part may be used for discrete information such as size/cross-section classification of aircraft radar images, wheeled/tracked classification of ground vehicles, etc.

Formally a data set is a random element $((y_i)^{N_M}_{i=1}, N_M, t, s)$ which is an element of

$$\bigcup_{m=0}^{\infty} \bigcup_{s \in S} \left((Y_s)^m \times \{m\} \times [t_0, \infty) \times \{s\} \right)$$

and represents N_M measurements, $y_1, \ldots, y_{N_M}$, generated by sensor s at time t. $(Y_s)^m = Y_s \times \ldots \times Y_s$ (m *times*) and $(Y_s)^0 = \{\theta\}$ where θ is a symbol for "no measurement." As we did for targets, we have included the number N_M of measurements in each data set to model objects with variable number of members explicitly. The purpose of the inclusion of (t, s) is to make clear the source of each information. Then we call any collection of such data sets a *cumulative data set*. The following six assumptions are used to further specify our generic sensor model:

---

[7] See the last remark in this section.

[8] synonymous to scans, frames, return sets, etc.

Assumption 3: (Finite Sampling Rate and Known Exact Timing) Every sensor generates at most one data set at any given time $t \in [t_0,\infty)$ and it produces only a finite number of data sets in any given finite interval in $[t_0,\infty)$. The time ($t \in [t_0,\infty)$) and the source ($s \in S$) are exactly known. The sensor scheduling is completely determined by each sensor and is assumed to be independent of the target system.

Assumption 4: (Conditional Independence) Given the target system stochastic process, each data set is independent from other data sets and depends only on the target system state at the time it is generated.

Assumption 5: (No Merged Measurement) Each measurement in each data set may or may not originate from a target. When a measurement in a data set originates from a target, it does not originate from two or more targets.

Assumption 6: (No Split Measurement) No target generates more than one measurement in any data set.

Assumption 7: (Random Order) The order of the measurements in any data set contains no information about targets.

Assumption 8: (Absolutely Continuous Distribution) For each data set from sensor s in S, given the target system state, the set of detected targets, the number of measurements and the assignment of the detected targets to the measurements, the joint distribution of measurement values has density with respect to hybrid measure μ_s.

Assumption 1 implies there are at most countable data sets generated in $[t_0,\infty)$. Thus, throughout the rest of this paper, we assume that the data sets are indexed by positive integers as[9], $z(1), z(2), z(3), \ldots\ldots$, where

$$z(k) = ((y_j(k))_{j=1}^{N_M(k)}, N_M(k), t_k, s_k) \tag{7}$$

is the k-th data set, in such a way that $t_k \leq t_{k'}$ whenever $k \leq k'$. We will call each $z(k)$ simply data set k. Since (t_k, s_k) is assumed to be exactly known (*Assumption* 3), we may treat them as non-

[9] Alternatively, we can index data sets by their third and fourth components, i.e., time and source (t,s), which by *Assumption 3* can identify data sets uniquely. This alternative indexing is useful when we consider target system state estimation based on different cumulative data sets.

random elements. *Assumption 4* then should be interpreted as follows: For any cumulative data set $\{z(1),.....z(K)\}$, we have

$$\begin{aligned} &Prob.\left(\bigcap_{k=1}^{K}\{z(k)\in(E_k\times\{(t_k,s_k)\})\}\,\middle|\,(X(t_k))_{k=1}^{K},N_T\right)= \\ &\qquad\prod_{k=1}^{K} Prob.\left(\{z(k)\in(E_k\times\{(t_k s_k)\})\}\,\middle|\,X(t_k),N_T\right) \end{aligned} \tag{8}$$

where each E_k is a measurable set in $\bigcup_{m=0}^{\infty}(Y_{s_k})^m\times\{m\}$, i.e., $E_k=\bigcup_{m=0}^{\infty}E_k^m\times\{m\}$ with each E_k^m being a measurable set in $(Y_{s_k})^m$. Therefore, modeling sensors means specifying each factor on the right hand side of (8) for each k.

One of unique notions involved in multitarget tracking problems is "detection of targets" or "origins of measurements," which can be modeled under *Assumptions 5 and 6* as follows: First we define a random set

$$I_T=\{i\in\mathbb{Z}_+\,|\,1\le i\le N_T\} \tag{9}$$[10]

which we call *target index set*. For each k, random set

$$J_M(k)=\{j\in\mathbb{Z}_+\,|\,1\le j\le N_M(k)\} \tag{10}$$

is called *measurement index set at k*. Then, for each k, we assume a random function[11] A_k defined on a subset of I_T and taking values in $J_M(k)$. We call such A_k *assignment function* at k. $j=A_k(i)$ means that the i-th target is *detected* by sensor s_k at time t_k and generates the j-th measurement, or j-th measurement *originates from* i-th target. Thus

$$I_{DT}(k)=Dom(A_k) \tag{11}$$[12]

is the random set consisting of indices for targets detected by sensor s_k at time t_k (or detected in data set k) and

[10] $\mathbb{Z}_+$ is the set of all the nonnegative integers.

[11] Usually a random function is defined as an random element whose samples share the same domain. But the domain of the random function A_k may differ from sample to sample.

[12] For a function f, $Dom(f)$ and $Im(f)$ are the domain and the range of f.

$$J_{FA}(k) = J_M(k) \backslash Im(A_k) \tag{12}$$[13]

is the set of indices of measurements in data set k which *do not originate from any target*. We call such measurements *false alarms*[14]. *Assumption 6* then implies that any realization of assignment function A_k is one-to-one. For the later use, we define random integers, i.e, the number of detected targets at k

$$N_{DT}(k) = \#(I_{DT}(k)) \tag{13}$$

and the number of false alarms in data set k

$$N_{FA}(k) = \#(J_{FA}(k)) \ . \tag{14}$$

Then we have an obvious relationship

$$N_M(k) = N_{DT}(k) + N_{FA}(k) \tag{15}$$

for each k.

We may visualize our generic sensor measurement generating mechanism as the following four-step processes, for each k:

(1) Detection: For each nonnegative integer n, let $\mathbf{D}(n)$ be the collection of all the subset of $\{1,.....,n\}$. $\mathbf{D}(n)=\{\varnothing\}$ when $n=0$. Apparently, $N_T=n$ implies $I_{DT}(k)\in\mathbf{D}(n)$ for each k. Then the detection mechanism can be specified by the *detection probability function* defined by

$$P_D(D \mid X, n, k) = Prob.\{I_{DT}(k)=D \mid X(t_k)=X, N_T=n\} \tag{16}$$

for each $(D, X, n)\in\bigcup_{n=0}^{\infty}\mathbf{D}(n)\times\mathbf{X}_n\times\{n\}$ and each k. When $D\notin\mathbf{D}(n)$, the right hand side of (16) is 0.

---

[13] $\backslash$ is the set subtraction operation, i.e., for two sets, A and B, $A\backslash B=\{a\in A \mid a\notin B\}$.

[14] They have also many other names such as false returns, clutters (in case of radars), nuisance targets (not real targets), etc.

$$J_{FA}(k) = J_M(k) \backslash Im(A_k) \qquad (12)^{13}$$

is the set of indices of measurements in data set k which *do not originate from any target*. We call such measurements *false alarms*[14]. *Assumption* 6 then implies that any realization of assignment function A_k is one-to-one. For the later use, we define random integers, i.e, the number of detected targets at k

$$N_{DT}(k) = \#(I_{DT}(k)) \qquad (13)$$

and the number of false alarms in data set k

$$N_{FA}(k) = \#(J_{FA}(k)) \ . \qquad (14)$$

Then we have an obvious relationship

$$N_M(k) = N_{DT}(k) + N_{FA}(k) \qquad (15)$$

for each k.

We may visualize our generic sensor measurement generating mechanism as the following four-step processes, for each k:

(1) Detection: For each nonnegative integer n, let $\mathbf{D}(n)$ be the collection of all the subset of $\{1,.....,n\}$. $\mathbf{D}(n)=\{\varnothing\}$ when $n=0$. Apparently, $N_T=n$ implies $I_{DT}(k)\in\mathbf{D}(n)$ for each k. Then the detection mechanism can be specified by the *detection probability function* defined by

$$P_D(D \mid X, n, k) = Prob.\{I_{DT}(k)=D \mid X(t_k)=X, N_T=n\} \qquad (16)$$

for each $(D, X, n)\in\bigcup_{n=0}^{\infty}\mathbf{D}(n)\times\mathbf{X}_n\times\{n\}$ and each k. When $D\notin\mathbf{D}(n)$, the right hand side of (16) is 0.

---

[13] $\backslash$ is the set subtraction operation, i.e., for two sets, A and B, $A\backslash B=\{a\in A \mid a\notin B\}$.

[14] They have also many other names such as false returns, clutters (in case of radars), nuisance targets (not real targets), etc.

(2) Number of Measurements: Because of relationship (15), given the number $N_{DT}(k)$ of detected targets in data set k, the number $N_M(k)$ of measurements is determined by the number $N_{FA}(k)$ of false alarms. Its statistics can be specified by the *number-of-false alarm probability function*,

$$P_{N_{FA}}(m \mid D, X, n, k) = Prob.\{N_{FA}(k)=m \mid I_{DT}(k)=D, X(t_k)=X, N_T=n\} \tag{17}$$

for every $(m, D, X, n) \in \mathbb{Z}_+ \times \left(\bigcup_{n=0}^{\infty} \mathbf{D}(n) \times \mathbf{X}_n \times \{n\}\right)$ and for each k. When $D \notin \mathbf{D}(n)$, the right hand side of (17) is arbitrary.

(3) Random Assignment: According to *Assumption 7*, given the set $I_{DT}(k)$ of detected targets and the number $N_M(k)$ of measurements for each data set k, the assignment of the origin of each measurement is totally random. In other words, every possible realization of assignment A_k is equally possible. For any pair (I, J) of finite sets, define

$$\mathbf{A}(I, J) = \{a : I \to J \mid a \text{ is one-to-one.}\} . \tag{18}$$

Then $\#(\mathbf{A}(I, J))$ is $\dfrac{\#(J)!}{(\#(J)-\#(I))!}$ when $\#(J) \geq \#(I)$ and 0 otherwise. Therefore we have, for any k,

$$Prob.\{A_k=a \mid N_M(k)=m, I_{DT}(k)=D, X(t_k)=X, N_T=n\} = \frac{(m-\#(D))!}{m!} \tag{19}$$

for every $(D, X, n) \in \bigcup_{n=0}^{\infty} \mathbf{D}(n) \times \mathbf{X}_n \times \{n\}$, every $m \geq \#(D)$ and every $a \in \mathbf{A}(D, \{1, \ldots, m\})$. When $a \notin \mathbf{A}(D, \{1, \ldots, m\})$, the left hand side of (19) is of course 0. It is arbitrary if $m < \#(D)$.

(4) Measurement Values: Through the above three steps, we have determined which targets are detected, the number of measurements and the origin of each measurement. The last step is to assign some measurement value to each measurement index $j \in J_M(k)$ being conditioned by $(X(t_k), N_T)$, $N_M(k)$ and A_k. By *Assumption 8*, we specify this step by the *measurement value probability density function* defined by

$$P_M(y \mid a, m, X, n, k)\mu_{s_k}^m(dy) = Prob.\{y \in dy \mid A_k=a, N_M(k)=m, X(t_k)=X, N_T=n\} \tag{20}$$

for every $(X, n) \in \bigcup_{n=1}^{\infty} \mathbf{X}_n \times \{n\}$, every (a, m) such that $a \in \mathbf{A}(D, \{1, \ldots, m\})$ and $m \geq \#(D)$ for some

$D \subseteq \{1,....,n\}$, and every $y \in (Y_{s_k})^m$, where $\mu_{s_k}^m$ is the m-tuple direct product measure of μ_{s_k}.

Thus our generic sensor model is specified by detection probability function P_D, number-of-false-alarm probability function $P_{N_{FA}}$ and measurement value probability density function P_M.

Remark 2: We should note that that the above three functions defining the generic sensor model in (16), (17) and (20) are, in general, all conditioned by the target system state at time t_k when each data set k is generated. In particular, (16) is not the joint probability of target detection but is a conditional probability of target detection conditioned by the target system state. Therefore, no-measurement data set $(\theta,0,t_k,s_k)$, i.e., "sensor s_k is up at time t_k but does not observe anything," is at least potentially as informative as other data sets with positive numbers of measurements. For example, for a sensor monitoring radio communication of a target, the probability of detecting the target is zero if the radio equipment of the target is turned off. Thus the absence of any reception while the sensor is up may provide vital information on the state of a target in some cases. In another example, an MTI-type[15] radar may detect targets only when the absolute values of targets' radial velocities exceed a given threshold. In such a case, the absence of radar return around the position where a target is expected to exist may be a valuable information indicating the targets stops somewhere in the vicinity of that position. It should also be noted that, in our formulation, every data set contains *number-of-measurement* information, and hence, every sensor is a *type-1* sensor in Reid's terminology in [4]. A *type-2* sensor in his terminology is a sensor which generates data sets each of which contains at most one measurement, i.e., $N_M(k) \leq 1$ for all k, and is not treated separately in our formulation.

Remark 3: For sensor systems with measurement time delays dependent on the target state (e.g., acoustic sensor systems described in [9]), a simple model where the target state is a pair, (position,velocity), may violate *Assumptions 3 and 4*, and/or other assumptions. In such a case, in order for our formulation to be applicable, careful modeling is necessary so that all the assumptions made so far are valid at least in an appropriate sense of approximation. Although *Assumptions 5 and 6* are quite standard in the multitarget tracking literature, their relaxation may be required in some applications. *Assumption 5* may be relaxed by assuming a random collection of subsets of target index set I_T and by considering each of such subsets as a potential origin for a measurement for each data set. Recently an algorithm which allows measurements to be merged has been developed in [10]. At this moment, however, it is not clear how to relax *Assumption 6*. In this report, we will not attempt to relax these assumptions. Another way to state *Assumption 7* is that a data set is the unit of sensor data in which the order of measurements in it does not contain any information about the targets. In some cases, however, the order of measurements from a set of returns may contain information about targets; Such may

[15] MTI = Moving Target Indicator

be a case when a radar uses a particular scanning pattern. In such a case, the data set should be further divided to the point where the measurement order does contain any significant information about targets.

Finally, in order to reflect the random-set nature of targets correctly, we must require the following interchangeability assumption in addition to *Assumption 2*.

Assumption 8: (Target Interchangeability (2)) Detection Probability function P_D, number-of-false-alarm probability function $P_{N_{FA}}$ and measurement value probability density function P_M are all invariant under target permutation, i.e., $P_D(\pi(D)\,|\,\Pi(X),n)$, $P_{N_{FA}}(m\,|\,\pi(D),\Pi(X),n)$ and $P_M(y\,|\,a_\pi,m,\Pi(X),n)$ are all invariant with respect to any n-target permutation homeomorphism Π induced by a permutation π on $\{1,\ldots,n\}$, where

$$\pi(D) = \{\pi(i)\,|\,i \in D\}$$

and, when $a \in \mathbf{A}(I,J)$, $a_\pi : \pi(Dom(a)) \rightarrow J$ is defined by[16]

$$a_\pi(i) = (a \circ \pi^{-1})(i) = a(\pi^{-1}(i))$$

for all $i \in \pi(Dom(a))$.

Remark 4: We may characterize the two target interchangeability assumptions, *Assumptions 2* and *8*, as *target without a priori identification*, although the use of term, "identification," may well be controversial since it is used in other areas of control theory with very distinctive meaning such as system identification, parameter identification, identification of weighting functions, etc. Suppose that the total number of targets is n and that there are m detected targets (of course $m \leq n$). Then, with the interchangeability assumptions, any one of $n!/(n-m)!$ possible ways of associating the m sets of data to the n targets is equally likely. This means that the targets are not discriminated from each other *a priori* or they are not *identified a priori*. On the other hand, if each target index has a specific meaning identifying a specific target, we may say the targets have *a priori* identification. This concept of target identification should also not be confused by "target classification." Almost all the "real" targets have unique identification, such as name, number plates, manufactory serial numbers, etc. However, in a target model where the targets do not have *a priori* identification, such identification is not considered within the model. Therefore the targets are labeled or discriminated with each other only through measurement indices. In such a model, the target classification is to determine the targets' discriminants as much as possible and the ultimate classification is to determine the

[16] $\circ$ is the function composition operator and π^{-1} is the inverse function of π.

identification of each target by means of a name, a registered number, etc., whenever it is possible. We will see more discussions on this subject in the next section as well as in a part of Section 6.

3. TRACKS AND HYPOTHESES

Although tracks and hypotheses are among most frequently used terms in the multitarget tracking literature, they are not often explicitly defined. Our definitions of tracks and hypotheses introduced in this section closely follow Morefield's notation in [11] but differ in one crucial aspect, namely separation of the measurement-value information from the number-of-measurement information in each data set. In other words, we will form tracks and hypotheses on the set of measurement indices rather than directly on the collection of data sets.

First we will remove the measurement-value information: For the moment, let Z be the cumulative set up to K, i.e.,

$$Z = \{z(k) \,|\, 1 \le k \le K\} \tag{21}$$

Then we call

$$\begin{aligned} J &= \bigcup_{k=1}^{K} J_M(k) \times \{k\} \\ &= \bigcup_{k=1}^{K} \{1, \ldots, N_M(k)\} \times \{k\} \end{aligned} \tag{22}$$

the *cumulative measurement index set* at K (or up to K or associating the cumulative data set Z). Then the cumulative data set Z represents all the information available up to and including the most recent data set $z(K)$. Every (j,k) in the cumulative measurement index set J represents the j-th measurement in data set k generated by sensor s_k at time t_k. Define a random collection of subsets of J by

$$\begin{aligned} \Lambda &= \left\{ \{(j,k) \,|\, j = A_k(i) \text{ and } 1 \le k \le K\} \;\middle|\; 1 \le i \le N_T \right\} \setminus \{\varnothing\} \\ &= \left\{ T_k(i) \triangleq \{(j,k) \,|\, j = A_k(i) \text{ and } 1 \le k \le K\} \neq \varnothing \;\middle|\; 1 \le i \le N_T \right\} . \end{aligned} \tag{23}$$

Each member of Λ is a set of measurement indices which are included in the cumulative measurement index set J and represent all the measurements originating from a target indexed by an $i \in I_T$. Since each A_k is a random one-to-one function (corresponding to the no-split/merged-measurement assumption, i.e., *Assumptions 5* and *6*), it follows from definition (23) that Λ must satisfies the following conditions:

(i) Each member of Λ is not empty and contains at most one measurement index for each data set.

(ii) Any two members of Λ do not share a common measurement index at any data set.

Then we call any possible realization λ of Λ a *data-to-data association hypothesis* or *data-to-data hypothesis* or simply *hypothesis* at K or on the cumulative data set Z. Let the set of all the *data-to-data association hypotheses* at K be denoted by $\mathbf{H}(Z)$. On the other hand, any subset τ of cumulative measurement index set J is a possible "trace" of a target (detected in at least one data set up to and including K) and called a *track* at K or on cumulative data set Z if it contains at most one measurement for each data set up to and including K. The set of all the tracks on cumulative data set Z (or at K) is denoted by $\mathbf{T}(Z)$. Namely, we have

$$\mathbf{T}(Z) = \{\tau \subseteq J \mid \#(\tau|k) \leq 1 \text{ for } \textit{all } k \textit{ such that } 1 \leq k \leq K\} \tag{24}$$[17]

and

$$\mathbf{H}(Z) = \left\{\lambda \subseteq \mathbf{T}(Z) \setminus \{\varnothing\} \;\middle|\; \begin{matrix} \tau_1 \cap \tau_2 = \varnothing \text{ for } \textit{all} \\ (\tau_1, \tau_2) \in \lambda \times \lambda \textit{ such that } \tau_1 \neq \tau_2 \end{matrix}\right\} . \tag{25}$$

In (24) and other subsequent equations, we define

$$\tau|k = \{j \in \mathbb{Z}_+ \mid (j,k) \in \tau\} \tag{26}$$

for any $\tau \subseteq \mathbb{Z}_+ \times \mathbb{Z}_+$. We should note that the empty collection $\varnothing$ is included in $\mathbf{H}(Z)$. We call such hypothesis *null hypothesis*, which hypothesizes that all the measurements in cumulative data set $Z^{(K)}$ are false alarms. Also $\mathbf{T}(Z)$ includes the empty set $\varnothing$ which we call *null track* and hypothesizes a target not detected in any of data sets up to and including K.

Given a cumulative data set Z, each $\lambda \in \mathbf{H}(Z)$ hypothesizes the following events:

(1) There are $\#(\lambda)$ targets which have been detected and generated at least one measurement in at least one data set up to and including K. This means number N_T of targets is at least $\#(\lambda)$.

(2) Each track τ in λ corresponds uniquely to a target which have been detected at least once up to and including to K.

[17] The definitions (24) (or (25), respectively) can be easily extended to the set of all the hypotheses (or tracks, resp.) on an arbitrary cumulative data set rather than Z defined by (21).

(3) $(j,k)\in\tau$ for a track τ in λ means that the j-th measurement in data set k originates from a target identified by track τ.

(4) For any $k\leq K$ and $\tau\in\lambda$, $\tau|k=\varnothing$ means that the target identified by τ is not detected (falsely dismissed) and hence generates no measurement in data set k.

(5) $J\setminus(\bigcup\lambda)$ is the set of all the measurement indices of false alarms in all the data sets up to and including to K.

It follows from the above discussions and the definitions Λ that

$$\begin{aligned} Prob.\{\Lambda\in\mathbf{H}(Z)\,|\,Z\} &= Prob.\{\Lambda\in\mathbf{H}(Z)\,|\,J\} \\ &= \sum_{\lambda\in\mathbf{H}(Z)} Prob.\{\Lambda=\lambda\,|\,Z\} \\ &= \sum_{\lambda\in\mathbf{H}(Z)} Prob.\{\Lambda=\lambda\,|\,J\} \\ &= 1 \end{aligned} \tag{27}$$

which means that set $\mathbf{H}(Z)$ of all the hypotheses on cumulative data set Z is a mutually distinct and collectively exhaustive set of "explanations" of the origins of all the measurements in cumulative data set Z. We should note that, in order to generate all the hypotheses in $\mathbf{H}(Z)$, we only need cumulative measurement index set J or equivalently *number-of-measurement sequence*, $N_M(1),N_M(2),\ldots\ldots$ The objective of the rest of this report to evaluate each hypothesis, i.e., to calculate its *a posteriori* probability $Prob.\{\Lambda=\lambda\,|\,Z\}$ and to infer the target system state $(X(t_k),N_T)$ under each hypothesis. We need a few preparations before the next section in which the main result of this report is stated.

First define a random subset of target index set I_T by

$$I = \bigcup_{k=1}^{K} Dom(A_k) = \bigcup_{k=1}^{K} I_{DT}(k) \tag{28}$$

which is the set of indices for all the targets detected at least once in a data set up to and including K and is called *cumulative detected target index set*. Then definition (23) of Λ implies $\#(I)=\#(\Lambda)$. Although Λ "explains" all the origins of measurements in cumulative data set Z, it does not tell us which track comes from which target. In other words, the target labeling by $\{1,..,N_T\}$ is lost in the set operation in (23). This uncertainty can be represented by another random function Ω which is defined on Λ and taking values in I_T. For each τ in Λ, $\Omega(\tau)$ is defined as an i such that

$$\tau = \{(j,k)\in J\ |\ j=A_k(i) \text{ and } 1\leq k\leq K\} \tag{29}$$

Since each A_k is a random one-to-one function, such an i is unique, and hence, Ω is well defined. Moreover any realization of Ω is also one-to-one and $Im(\Omega)=I$. We call any realization of Ω a *target-to-track hypothesis* on Z and any realization (ω,λ) of (Ω,Λ) a *target-to-data hypothesis* on Z. Let $\mathbf{W}(\lambda,n)$ be the collection of all the target-to-track hypotheses under assumptions $\Lambda=\lambda$ and $N_T=n$, i.e.,

$$\mathbf{W}(\lambda,n)=\left\{\omega:\lambda\rightarrow\{1,...,n\}\,|\,\omega\ is\ one-to-one.\right\}\ . \tag{30}$$

We should note that pair (Ω,Λ) is nothing but another form of representation for the assignment tuple $(A_k)_{k=1}^K$. This is so since, for each k, we have

$$Dom(A_k)=I_{DT}(k)=\left\{i\in Im(\Omega)\ \middle|\ \Omega^{-1}(i)\,|k\neq\varnothing\right\} \tag{31}$$

and

$$j=A_k(i)\quad \text{if and } only \text{ if}\quad \{j\}=\Omega^{-1}(i)\,|k\ . \tag{32}$$

As a consequence of our target interchangeability conditions, we have the following important lemma:

Lemma 1: For each $\lambda\in\mathbf{H}(Z)$, for each $n\geq\#(\lambda)$ and each $\omega\in\mathbf{W}(\lambda,n)$, we have

$$Prob.\{\Omega=\omega\,|\,\Lambda=\lambda,N_T=n\}=\#(\mathbf{W}(\lambda,n))^{-1}=\frac{(n-\#(\lambda))!}{n!}\ . \tag{33}$$

Proof: Since $\mathbf{W}(\lambda,n)$ contains all the target-to-track hypotheses when conditioned by $\Lambda=\lambda$ and $N_T=n$, we have

$$\sum_{\omega\in\mathbf{W}(\lambda,n)}Prob.\{\Omega=\omega\,|\,\Lambda=\lambda,N_T=n\}=1\ . \tag{34}$$

On the other hand, if $\omega_1\in\mathbf{W}(\lambda,n)$ and $\omega_2\in\mathbf{W}(\lambda,n)$, there exists a permutation π on $\{1,..,n\}$ such that $\omega_1=_{\mathbf{S}}\omega_2$ or $\omega_1(\tau)=\pi(\omega_2(\tau))$ for all $\tau\in\lambda$. Since $\#(\mathbf{W}(\lambda,n))=\frac{n!}{(n-\#(\lambda))!}$, to show (34), it suffices to show that $Prob.\{\Omega=_{\mathbf{S}}\omega\,|\,\Lambda=\lambda,N_T\}$ or equivalently $Prob.\{\Omega=_{\mathbf{S}}\omega,\Lambda=\lambda\,|N_T\}$ is invariant under any permutation π on $\{1,...,n\}$. Since (Ω,Λ) is one-to-one to $(A_k)_{k=1}^K$, there exists $(a_k)_{k=1}^K$ such that

$$\{\Omega=\omega,\Lambda=\lambda\} = \bigcap_{k=1}^{K} \{A_k=a_k\}$$

Thus, using *Assumption 4 (Conditional Independence)*, we have

$$\begin{aligned}
&Prob.\ \{\Omega=\omega,\Lambda=\lambda \mid N_T=n\} \\
&= \int_{\mathbf{X}_n} Prob.\ \{\Omega=\omega,\Lambda=\lambda \mid (X(t_k))_{k=1}^{K}=(X_k)_{k=1}^{K},N_T=n\}\ Prob.\ \{(X(t_k))_{k=1}^{K}\in\prod_{k=1}^{K} dX_k \mid N_T=n\} \\
&= \int_{\mathbf{X}_n} Prob.\ (\bigcap_{k=1}^{K}\{A_k=a_k\} \mid (X(t_k))_{k=1}^{K}=(X_k)_{k=1}^{K},N_T=n)\ Prob.\ \{(X(t_k))_{k=1}^{K}\in\prod_{k=1}^{K} dX_k \mid N_T=n\} \\
&= \int_{\mathbf{X}_n} \prod_{k=1}^{K} Prob.\ \{A_k=a_k \mid X(t_k)=X_k,N_T=n\}\ Prob.\ \{(X(t_k))_{k=1}^{K}\in\prod_{k=1}^{K} dX_k \mid N_T=n\}
\end{aligned}$$

Using (16), (17) and (19), we have, for each k,

$$\begin{aligned}
&Prob.\ \{A_k=a_k \mid X(t_k)=X_k,N_T=n\} = P_D(Dom(a_k) \mid X_k,n) \\
&\left(\sum_{m=max\{Im(a_k)\}-\#(Im(a_k))}^{\infty} \frac{m!}{(m+\#(Im(a_k)))!} P_{N_{FA}}(m \mid Dom(a_k),X_k,n)\right)
\end{aligned}$$

Therefore, it follows from the target interchangeable assumptions, *Assumptions 2 and 9*, that $Prob.\ \{\Omega=_{\pi}\omega,\Lambda=\lambda \mid N_T=n\}$ is invariant with respect to π. *Q.E.D.*

This lemma is very important. First this lemma states that we do not have to evaluate every target-to-data hypotheses (ω,λ) instead evaluation of every *data-to-data hypothesis* is sufficient. Secondly this lemma gives a hint as to what should be the appropriate set of variables to propagate forward when we evaluate each hypothesis recursively. The importance of this choice may be illustrated in the following very simple example:

Example 1: Assume $Prob.\ \{N_T=2\}=1$, $\mathbf{X}_2=\{S_1,S_2\}$, and targets are stationary and i.i.d., i.e.,

$$Q_0^2(E) = \sum_{(i_1,i_2)\in E} q(i_1)q(i_2)$$

with some $q : \mathbf{X}_2 \rightarrow (0,1)$ such that $q(S_1)+q(S_2)=1$. Suppose we have a data set

$z(1)=(S_1,S_2,t_1,s_1)$ from a sensor s_1 at time $t_1 \geq t_0$ with $P_D(\{1,2\}|X,2,1)=1$, $P_{N_{FA}}(0|D,X,2,1)=1$, and $P_M(X_{a(1)},X_{a(2)}|a,2,(X_1,X_2),2,1)=1$, where $Y_s=\{S_1,S_2\}$. Namely, the probability detection is one, no possibility of false alarms and the measurement error is zero. Let $Z=\{z(1)\}$. Then apparently *Prob.* $\{\Lambda=\lambda|Z\}=1$ where $\lambda=\{\tau_1,\tau_2\}$, $\tau_1=\{(1,1)\}$ and $\tau_2=\{(2,1)\}$. Then, by *Lemma 1*, *Prob.* $\{\Omega=\omega|\Lambda=\lambda,N_T=2\}=1/2$ for each one-to-one function ω from λ to $\{1,2\}$. Thus, we have

$$Prob.\{X(t_1)=(S_1,S_2)|N_T=2,\Lambda=\lambda,Z\}=Prob.\{X(t_1)=(S_2,S_1)|N_T=2,\Lambda=\lambda,Z\}=1/2.$$

Namely conditional distribution of targets is no longer independent although the *a priori* distribution is independent. This is a direct consequence of our assumption of no merged measurement, i.e, if a target is at S_1 the other target must be at S_2, and vice versa. The mixing of the target state distribution with all the realizations of Ω, which are all equally likely because of the target interchangeability assumptions, thus turns the independent *a priori* distribution into *a posteriori* distribution with cross-correlation. In this particular example, the *a posteriori* distribution is still interchangeable. In general, however, this kind of mixture may create very complicated *a posteriori* distributions. On the other hand, if we conditioned the state distribution by one of the equally probable realization of Ω, then the *a posteriori* distribution is still independent although, in this particular example, the distribution is degenerated.

This example suggests that

$$\begin{aligned} &Prob.\{X(t_K)\in dX\,|N_T=n,\Lambda=\lambda,Z\} \\ &=\sum_{\omega\in W(\lambda,n)} Prob.\{X(t_K)\in dX\,|N_T=n,\Omega=\omega,\Lambda=\lambda,Z\}\,Prob.\{\Omega=\omega|N_T=n,\Lambda=\lambda,Z\} \\ &=\frac{(n-\#(\lambda))!}{n!}\sum_{\omega\in W(\lambda,n)} Prob.\{X(t_K)\in dX\,|\Omega=\omega,N_T=n,\Lambda=\lambda,Z\} \end{aligned}$$

may not be in an appropriate set of variables to propagate forward when we evaluate hypotheses recursively and infer the target state (X,N_T) accordingly. We may say the above mixture is an unnecessary "over-aggregation" of information. For this reason,

$$Prob.\{X(t_K)\in dX\,|\Omega=\omega,\Lambda=\lambda,N_T,Z\}$$

is more appropriate as shown in the next section.

Before we close this section, we will investigate the relationship among tracks and hypotheses. For this purpose, we consider all the cumulative data sets and, to explicitly

represent the difference, we denote the cumulative data set up to K by $Z^{(K)}$. The corresponding cumulative measurement index sets are denoteed by $J^{(K)}$. Then such cumulative sets (and cumulative measurement index sets) are finitely or at most countably many. Then apparently we have the inclusion relationship,

$$\mathbf{T}(Z^{(k_1)}) \subseteq \mathbf{T}(Z^{(k_2)}) \qquad \text{and} \qquad \mathbf{H}(Z^{(k_2)}) \subseteq \mathbf{T}(Z^{(k_2)})$$

for all $k_1 \leq k_2$. Redefine Z as the maximum cumulative data sets, i.e., $\bigcup_{k\geq 1} z(k) = \bigcup_{k\geq 1} Z^{(k)}$. Then, for any track $\tau \in \mathbf{T}(Z)$, we call $\tau \cap J^{(k)}$ the *restriction of track* τ *on* $Z^{(k)}$. For any hypothesis $\lambda \in \mathbf{H}(Z)$, we call

$$\lambda \,|\, Z^{(k)} = \{\tau \cap J^{(k)} \,|\, \tau \in \lambda\} \setminus \{\varnothing\}$$

the *restriction hypothesis* λ *on* $Z^{(k)}$. When a track τ_1 is a restriction of τ_2, we say τ_1 is a *predecessor* of τ_2, or equivalently, τ_2 is a *successor* of τ_1. This predecessor/successor relationship is defined on $\mathbf{H}(Z)$ through the above restriction operation in the same way. Then each predecessor/successor relationship defined a partial order on each of $\mathbf{T}(Z)$ and $\mathbf{H}(Z)$.[18] An immediate predecessor of any track or any hypothesis is unique and called a *parent* (always single). Therefore both $\mathbf{T}(Z)$ and $\mathbf{H}(Z)$ are *arborescent* (*tree-like*) with such partial orderings, i.e., the set of predecessors of any element is totally ordered. This justifies the commonly used terminology such as *hypothesis tree* or *track splitting.*

We conclude this section by the following remark:

Remark 5: According to Morefield's notation in [11][19], a track is a subset of a cumulative data set Z rather than a cumulative measurement index set J as in our formulation. When a track is defined in such a way, it include measurement values, and hence, the *a priori* probability distribution or its density of a hypothesis (as defined as a possible collection of tracks) is very hard to calculate and may not be well defined. For example, the validity of the following two equations[20] is questionable:

$$P(\lambda) = \exp(-\nu_{FA})\,(\nu_{FA})^{N_{FA}}/N_{FA}!$$

where ν_{FA} is the expected number of false alarms and N_{FA} is the hypothesized number of false alarms, and

[18] The partial order on $\mathbf{T}(Z)$ is merely the set-inclusion partial order.

[19] In Morefield's notation, τ is a hypothesis and λ is a track.

[20] Each of these equation is in [11].

$$P(\lambda) = \prod_{\tau \in \lambda} exp(-l_\tau / L)$$

where l_τ is the length of track τ and L is the expected track length. We will present an alternative batch-processing type formulation in Section 6 and a possible application 0-1 linear programming techniques.

4. GENERAL RESULTS

This section describes the main result of this report, i.e., the recursive hypothesis evaluation formula. All the assumptions made in Section 2 will be used. However, we should note that we have made very general assumptions which allow dependence among targets although the target interchangeability is still crucial to our derivation of the general algorithm. To explicitly specify cumulative data sets, we denote cumulative data set up to k by $Z^{(k)}$, and accordingly, the cumulative measurement index set up to k by $J^{(k)}$. Instead of (23) and (29), we use

$$\Lambda_k = \left\{ \{(A_k(i),k)' | i \in I_{DT}(k)' \text{ and } 1 \le k \le k\} \,\middle|\, 1 \le i \le N_T \right\} \setminus \{\varnothing\}$$

and

$$\Omega_k(\tau) = i \quad \text{if and } \textit{only} \text{ if} \quad \tau = \{(j,k) \in J \mid j = A_k(i) \text{ and } 1 \le k \le k\} \ .$$

Each hypothesis on $Z^{(k)}$ can be evaluated recursively according to index k. With usual notational abuse of P, we have

$$P(\Lambda_k | Z^{(k)}) = \frac{P(Z^{(k)}, \Lambda_k | Z^{(k-1)}, \Lambda_{k-1})}{P(Z^{(k)} | Z^{(k-1)})} P(\Lambda_{k-1} | Z^{(k-1)}) \tag{35}$$

for each k. Since $P(Z^{(k)} | Z^{(k-1)})$ is the normalizing constant, the hypothesis evaluation, i.e., the calculation of $P(\Lambda_k | Z^{(k)})$, can be accomplished by calculating $P(Z^{(k)}, \Lambda_k | Z^{(k-1)}, \Lambda_{k-1})$ if $P(\Lambda_{k-1} | Z^{(k-1)})$ is known under the recursion assumption. This is done by considering all the factors involved in this term, i.e., the sensor models specified by the detection probability function P_D, the number-of-false-alarm probability function $P_{N_{FA}}$, the measurement-value

probability density function P_M, the target system state estimate based upon the previous cumulative data set $Z^{(k-1)}$, and at last most importantly, by random assignment of measurements and target interchangeability.

To accomplish the recursion, we must chose an appropriate set of variables[21]. We should have a complete set of information from which we can calculate $P(Z^{(k)},\Lambda_k \mid Z^{(k-1)},\Lambda_{k-1})$ while we wish to minimize the set of variables which we must carry forward. Taking the consideration made at the end of the last section into account, the following set of variables to be propagated is chosen: For each $k>0$, we define

$$P_{\mathbf{H}}^{(k)}(\lambda \mid Z^{(k)}) = Prob.\{\Lambda(k)=\lambda \mid Z^{(k)}\} \quad , \tag{36}$$

$$P_{N_T}^{(k)}(n \mid \lambda, Z^{(k)}) = Prob.\{N_T=n \mid \Lambda(k)=\lambda, Z^{(k)}\} \quad \text{, and} \tag{37}$$

$$P_{\mathbf{X}}^{(k)}(dX \mid \omega, n, \lambda, Z^{(k)}) = Prob.\{X(t_k)\in dX \mid \Omega_k=\omega, N_T=n, \Lambda(k)=\lambda, Z^{(k)}\}, \tag{38}$$

for each $\lambda\in\mathbf{H}(Z^{(k)})$, each $n\geq\#(\lambda)$[22], and *some* $\omega\in\mathbf{W}(\lambda,n)$. Only one ω is necessary because, in an appropriate sense, (38) is invariant with respect to any $\omega\in\mathbf{W}(\lambda,n)$, as shown in the following theorem. The evaluation of such an ω is also unnecessary thanks to *Lemma 1* in the previous section. The following theorem is the main result of this report. For this theorem, we define $Z^{(0)}=\varnothing$,

$$P_{\mathbf{H}}^{(0)}(\varnothing \mid \varnothing)=1 \quad , \tag{36'}$$

$$P_{N_T}^{(0)}(n \mid \varnothing, \varnothing) = Prob.\{N_T=n\} \quad \text{and} \tag{37'}$$

$$P_{\mathbf{X}}^{(0)}(dX \mid \varnothing, n, \varnothing, \varnothing) = Q_0^n(dX) \quad . \tag{38'}$$

Theorem 1: (General Result) For each $k>0$, let $Z=Z^{(k)}$ and $\bar{Z}=Z^{(k-1)}$. Then, for each $\lambda\in\mathbf{H}(Z)$, when $\bar{\lambda}$ is the unique parent of λ on $\bar{Z}$, we have

[1] *(Target System State Estimate Updating) for any $n\geq\#(\lambda)$ and any $\omega\in\mathbf{W}(\lambda,n)$,*

[21] The variables (or more appropriately called "functions") which are to propagate forward. They may be called *filtering states* or *sufficient statistics*.

[22] By the definition of Λ_k, $Prob.\{N_T<\#(\Lambda_k)\}=0$ for any k.

$$P_X^{(k)}(dX \mid \omega,n,\lambda,Z) = C_X^{(k)}(z(k) \mid \omega,\lambda,n,\bar{Z})^{-1}$$
$$P_M((y_j(k))_{j=1}^{N_M(k)} \mid a,N_M(k),X,n,k)\ P_{N_{FA}}(N_M(k) - n_D(\lambda \mid k) \mid D,X,n,k)$$
$$P_D(D \mid X,n,k) \int_{X_n} F^n_{t_k - t_{k-1}}(dX \mid \bar{X})\, P_X^{(k-1)}(d\bar{X} \mid \bar{\omega},n,\bar{\lambda},\bar{Z}) \tag{39}$$

where C_X *is the normalizing constant for (39) and defined by*

$$C_X^{(k)}((y,m,J_k,s_k) \mid \omega,\lambda,n,\bar{Z}) =$$
$$\int_{X_n} P_M(y \mid a,m,X,n,k)\ P_{N_{FA}}(m - n_D(\lambda \mid k) \mid D,X,n,k)$$
$$P_D(D \mid X,n,k) \int_{X_n} F^n_{t_k - t_{k-1}}(dX \mid \bar{X})\, P_X^{(k-1)}(d\bar{X} \mid \bar{\omega},n,\bar{\lambda},\bar{Z}) \tag{40}$$

for each $(y,m) \in \bigcup_{m=0}^{\infty} (Y_{s_k})^m \times \{m\}$, $n_D(\lambda \mid k)$ *is the number of targets which are hypothesized (by* λ*) to be detected in data set* $z(k)$*, i.e.,*

$$n_D(\lambda \mid k) = \sum_{\tau \in \lambda} \#(\tau \mid k) \quad , \tag{41}$$

and $\bar{\omega} \in W(\bar{\lambda},n)$, $D \in \mathbf{D}(n)$ *and* $a \in \mathbf{A}(D,J_M(k))$ *are uniquely defined by*

$$\omega(\tau) = \bar{\omega}(\tau \cap J^{(k-1)}) \qquad \text{for every } \tau \in \lambda \text{ such that } \tau \cap J^{(k-1)} \neq \emptyset \ , \tag{42}$$
$$D = Dom(a) = \{i \in Im(\omega) \mid (\omega^{-1}(i) \mid k) \neq \emptyset\} \quad \text{and} \tag{43}$$
$$\{a(i)\} = \omega^{-1}(i) \mid k \qquad \text{for all } i \in D \ , \tag{44}$$

[2] *(Target Interchangeability) for any* $n \geq \#(\lambda)$, $C_X^{(k)}(\cdot \mid \omega,\lambda,n,\bar{Z})$ *defined by* (40) *is invariant for all* $\omega \in W(\lambda,n)$*, and moreover, for any* $\omega \in W(\lambda,n)$ *and any permutation* π *on* $\{1,\ldots,n\}$*,*

$$P_X^{(k)}(dX \mid \omega,n,\lambda,Z) = P_X^{(k)}(\Pi(dX) \mid \omega,n,\lambda,Z) \tag{45}$$

if $\omega \doteq_{\mathbf{S}} \omega$ *and* $\Pi : X_n \to X_n$ *is the n-target permutation homeomorphism induced by* π*.*

[3] *(Number-of-Target Estimate Updating) for any* $n \geq \#(\lambda)$*,*

$$P_{N_T}^{(k)}(n\,|\lambda,Z) = \mathbf{C}_{N_T}^{(k)}(z(k)\,|\lambda,\bar{Z})^{-1}\,\mathbf{C}_{N}^{(k)}(z(k)\,|\omega,\lambda,n,\bar{Z})\frac{(n-\#(\bar{\lambda}))!}{(n-\#(\lambda))!}P_{N_T}^{(k-1)}(n\,|\bar{\lambda},\bar{Z}) \tag{46}$$

where $\mathbf{C}_{N_T}^{(k)}$ *is the normalizing constant defined by*

$$\mathbf{C}_{N_T}^{(k)}((y,m,s_k,s_k)\,|\lambda,\bar{Z}) = \sum_{n=\#(\lambda)}^{\infty}\mathbf{C}_{N}^{(k)}((y,m,s_k,s_k)\,|\omega,\lambda,n,\bar{Z})\,\frac{(n-\#(\bar{\lambda}))!}{(n-\#(\lambda))!}P_{N_T}^{(k-1)}(n\,|\bar{\lambda},\bar{Z}) \tag{47}$$

for each $(y,m)\in\bigcup_{m=0}^{\infty}(Y_{s_k})^m\times\{m\}$ *and* ω *is an arbitrary target-to-track hypothesis in* $\mathbf{W}(\lambda,n)$,

and

[4] *(Hypothesis Evaluation)*

$$P_{\mathbf{H}}^{(k)}(\lambda\,|Z) = \frac{\mathbf{L}_k(z(k),\lambda\,|\bar{Z},\bar{\lambda})}{P_{Z}^{(k)}(z(k)\,|\bar{Z})}\,P_{\mathbf{H}}^{(k-1)}(\bar{\lambda}\,|\bar{Z}) \tag{48}$$

where $P_{Z}^{(k)}(z(k)\,|\bar{Z})$ *is the normalizing constant and* $\mathbf{L}_k(z(k),\lambda\,|\bar{Z},\bar{\lambda})$ *is the likelihood of* $(z(k),\lambda)$ *given* $(\bar{Z},\bar{\lambda})$ *defined by*

$$\mathbf{L}_k((y,m,s_k,s_k),\lambda\,|\bar{\lambda},\bar{Z}) = \frac{(m-n_D(\lambda\,|k))!}{m!}\,\mathbf{C}_{N_T}^{(k)}((y,m,s_k,s_k)\,|\lambda,\bar{Z}) \tag{49}$$

for each $(y,m)\in\bigcup_{m=0}^{\infty}(Y_{s_k})^m\times\{m\}$ *with* $n_D(\lambda\,|k)$ *and* $\mathbf{C}_{N_T}^{(k)}$ *being defined by* (41) *and* (47), *respectively.*

Proof: First we will prove [1] and [2]: To prove [1], roughly speaking we follow the Bayesian expansion,

$$P(dX\,|\Omega_k,N_T,\Lambda_k,Z) = \frac{P(Z,\Lambda_k,\Omega_k\,|\Lambda_{k-1},\Omega_{k-1},X,N_T,\bar{Z})}{P(Z,\Lambda_k,\Omega_k\,|\Lambda_{k-1},\Omega_{k-1},N_T,\bar{Z})}P(dX\,|\Omega_{k-1},N_T,\Lambda_{k-1},\bar{Z})$$

abusing notation P. Therefore we must calculate $P(Z,\Lambda_k,\Omega_k\,|\Lambda_{k-1},\Omega_{k-1},X,N_T,\bar{Z})$. Let m and n be arbitrary nonnegative integers, E any measurable set in $(Y_k)^m$, $\lambda\in\mathbf{H}(Z)$ any data-to-data hypothesis on Z, $\omega\in\mathbf{W}(\lambda,n)$ be any target-to-track hypothesis compatible with (λ,n), $\bar{\lambda}\in\mathbf{H}(Z)$

any data-to-data hypothesis on $\bar{Z}$, and $\bar{\omega}\in W(\bar{\lambda},n)$ be any target-to-track hypothesis compatible with $(\bar{\lambda},n)$. Then a straightforward but rather lengthy application of Bayes rule gives us the following expansion:

$$\begin{aligned}
&Prob.\{z(k)\in E\times\{m,l_k,s_k\},\Lambda_k=\lambda,\Omega_k=\omega\,|\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad=\sum_{D\in D(n)} Prob.\{I_{DT}(k)=D\,|\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad Prob.\{N_M(k)=m\,|I_{DT}(k)=D,\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\sum_{a\in A(D,\{1,\ldots,m\})} Prob.\{A_k=a\,|N_M(k)=m,I_{DT}(k)=D,\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\} \qquad (50)^{23}\\
&\quad Prob.\{\Lambda_k=\lambda\,|A_k=a,N_M(k)=m,\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad Prob.\{\Omega_k=\omega\,|A_k=a,N_M(k)=m,\Lambda_k=\lambda,\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad Prob.\{(y_j(k))_{j=1}^m\in E\,|A_k=a,N_M(k)=m,\Lambda_k=\lambda,\Omega_k=\omega,X(t_k)=X,N_T=n,\bar{Z}\}
\end{aligned}$$

The first, the second, the third and the sixth factors on the right hand side of the above equation are given by (16), (17), (19) and (20), respectively. Namely, for the first factor, we have

$$\begin{aligned}
&Prob.\{I_{DT}=D\,|\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad=Prob.\{I_{DT}=D\,|X(t_k)=X,N_T=n\}=P_D(D\,|X,n,k)
\end{aligned}$$

since data sets are conditionally independent. For the same reason, the second factor is given by

$$\begin{aligned}
&Prob.\{N_M(k)=m\,|I_{DT}=D,\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}\\
&\quad=Prob.\{N_M(k)=m\,|I_{DT}=D,X(t_k)=X,N_T=n\}\\
&\quad=\begin{cases}P_{N_{FA}}(m-\#(D)\,|D,X,n,k) & \text{if } m\geq\#(D)\\ 0 & \textit{otherwise}\end{cases}
\end{aligned}$$

Since $I_{DT}(k)=Dom(A_k)$, the third factor is given by

$$Prob.\{A_k\,|N_M(k)=m,I_{DT}=D,\Lambda_{k-1}=\bar{\lambda},\Omega_k=\bar{\omega},X(t_k)=X,N_T=n,\bar{Z}\}$$

$$=Prob.\{A_k\,|N_M(k)=m,I_{DT}=D,X(t_k)=X,N_T=n\}=\begin{cases}\dfrac{(m-\#(D))!}{m!} & \text{if } Dom(a)=D\\ 0 & \textit{otherwise}\end{cases}$$

[23] (50) is composed as $P_0=\sum P_1P_2\sum P_3P_4P_5P_6$. By the i-th factor on the right hand side of (50), we mean conditional probability P_i.

In the sixth factor, as far as data set $z(k)$ is concerned, the condition by assignment hypothesis a and the condition by target-to-data hypothesis (ω,λ) are redundant. Thus we have

$$\begin{aligned} &Prob.\{(y_j(k))_{j=1}^{m}\in E \mid A_k=a, N_M(k)=m, \Lambda_k=\lambda, \Omega_k=\omega, X(t_k)=X, N_T=n, \bar{Z}\} \\ &= \int_E P_M((y_j(k))_{j=1}^{m} \mid a, m, X, n, k) \prod_{j=1}^{m} \mu_{s_k}(dy_j) \quad . \end{aligned}$$

On the other hand, the fourth and the fifth factors merely check the consistency among λ, $\bar{\lambda}$, ω, $\bar{\omega}$ and a. The fourth factor is 1 if $\bar{\lambda}$ is the unique parent of λ on $\bar{Z}$ and if

$$\lambda = \lambda_{old} \cup \lambda_{new} \tag{51}$$

where

$$\begin{aligned} \lambda_{old} &= \left\{\bar{\tau}\cup\{(j,k) \mid j=a(\bar{\omega}(\bar{\tau}))\} \;\middle|\; \bar{\tau}\in\bar{\lambda}\right\} \quad \text{and} \\ \lambda_{new} &= \left\{\{(a(i),k)\} \;\middle|\; i\in Dom(a)\setminus Im(\bar{\omega})\right\} \quad . \end{aligned} \tag{52}$$

It is 0 otherwise. In (51) and (52), λ_{old} *is the set of tracks which have been already started in* $J^{(k-1)}$ and λ_{new} is the set of tracks corresponding to the targets which λ hypothesizes to be newly detected in data set $z(k)$. The fifth factor is 1 if $\bar{\omega}$ is the restriction of ω to $\bar{\lambda}$ in the sense of (42) and the target-to-track assignment by ω is consistent with the target-to-measurement-index assignment by a, i.e., it is 1 if

$$\begin{cases} \omega(\tau) = \bar{\omega}(\tau\cap J^{(k-1)}) & \text{if } \tau\in\lambda_{old} \\ \tau\mid k = \{a(\omega(\tau))\} & \text{if } \tau\in\lambda_{new} \end{cases} \tag{53}$$

and 0 otherwise. Therefore, for a given target-to-data hypothesis (ω,λ), the fourth factor times the fifth factor is 1 if $(\bar{\omega},D,a)$ satisfies (42) to (44)[24] and 0 otherwise. Then, since a is one-to-one, (44) implies $n_D(\lambda\mid k)=\sum_{\tau\in\lambda}\#(\tau\mid k)=\#(Dom(a))$. Thus (50) is reduced to

[24] In other words, if $(\bar{\omega},D,a)$ is uniquely determined by (42) to (44) from given (ω,λ).

$$\begin{aligned} &Prob.\{\zeta(k)\in E\times\{m\},\Lambda_k=\lambda,\Omega_k=\omega\,|\,\Lambda_{k-1}=\bar{\lambda},\Omega_k=\bar{\omega},X(t_k)=X,N_T,\bar{Z}\} \\ &= P_D(D\,|X,n,k)\;P_{N_{FA}}(m-n_D(\lambda\,|k)\,|D,X,n,k) \\ &\qquad \frac{(m-n_D(\lambda\,|k))!}{m!}\int_E P_M((y_j)_{j=1}^m\,|a,n,X,n,k)\prod_{j=1}^m \mu_{s_k}(dy_j) \end{aligned} \tag{54}$$

if $m\geq\#(D)$ and $(\bar{\omega},D,a)$ satisfies (42) to (44). Otherwise the left hand side of (54) is zero. Equation (54) together with "extrapolation" equation,

$$\begin{aligned} &Prob.\{X(t_k)\in dX\,|\Omega_{k-1}=\bar{\omega},N_T=n,\Lambda_{k-1}=\bar{\lambda},\bar{Z}\} \\ &= \int_{\mathbf{X}_n} F^n_{t_k-t_{k-1}}(dX\,|\bar{X})P^{(k-1)}_{\mathbf{X}}(d\bar{X}\,|\bar{\omega},n,\bar{\lambda},\bar{Z}) \end{aligned} \tag{55}$$

implies (39) with (40), which concludes the proof for part [1]. Then part [2] follows immediately: The second statement of [2] for $k=0$, which should mean (5) in Section 2, is true by *Assumption 2*. Suppose the second statement of part [2] is true for $k-1$. Then , by the interchangeability assumptions, *Assumptions 2 and 9*, we have

$$\begin{aligned} &P_M((y,m)\,|\omega\,\pi^{-1},m,\Pi(X),n,k)\;P_{N_{FA}}(m-n_D(\lambda\,|k)\,|\pi(D),\Pi(X),n,k) \\ &\qquad P_D(\pi(D)\,|\Pi(X),n,k)\int_{\mathbf{X}_n} F^n_{t_k-t_{k-1}}(\Pi(dX)\,|\Pi(\bar{X}))\,P^{(k-1)}_{\mathbf{X}}(\Pi(d\bar{X})\,|\,\omega,n,\bar{\lambda},\bar{Z}) \end{aligned}$$

is invariant for any n-target permutation homeomorphism Π induced by any permutation π on $\{1,\ldots,n\}$. This means that the normalizing constant $\mathbf{C}^{(k)}_{\mathbf{X}}(\cdot\,|\omega,\lambda,n,\bar{Z})$ is also invariant for all $\omega\in\mathbf{W}(\lambda,n)$ (the first statement of part [2]) and the second statement of part [2] is also true for k. This completes the proof of parts [1] and [2].

Next we will prove [3]: To do this, roughly speaking, we must calculate $P(Z^{(k)},\Lambda_k\,|\Lambda_{k-1},N_T,Z^{(k-1)})$. Let m be any nonnegative integer, E any measurable set in $(Y_{s_k})^m$, $\lambda\in\mathbf{H}(Z)$ any data-to-data hypothesis on Z and n any nonnegative integer such that $n\geq\#(\lambda)$. Then it follows from a straightforward application of Bayes rule and from (40), (54)

and (55) that

$$
\begin{aligned}
&\text{Prob.}\{z(k)\in E\times\{m,s_k\},\Lambda_k=\lambda\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}\\
&=\sum_{\omega\in W(\bar{\lambda},n)}\text{Prob.}\{z(k)\in E\times\{m,s_k\},\Lambda_k=\lambda,\Omega_k=\omega\mid\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=(\omega\mid\bar{\lambda}),N_T=n,\bar{Z}\}\\
&\qquad\text{Prob.}\{\Omega_{k-1}=(\omega\mid\bar{\lambda})\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}\\
&=\sum_{\omega\in W(\bar{\lambda},n)}\text{Prob.}\{\Omega_{k-1}=(\omega\mid\bar{\lambda})\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}\\
&\qquad\int_X\text{Prob.}\{z(k)\in E\times\{m\},\Lambda_k=\lambda,\Omega_k=\omega\mid\Lambda_{k-1}=\bar{\lambda},\Omega_{k-1}=(\omega\mid\bar{\lambda}),X(t_k)=X,N_T=n,\bar{Z}\}\\
&\qquad\qquad\text{Prob.}\{X(t_k)\in dX\mid\Omega_{k-1}=(\omega\mid\bar{\lambda}),N_T=n,\Lambda_{k-1}=\bar{\lambda},\bar{Z}\}\\
&=\sum_{\omega\in W(\bar{\lambda},n)}\text{Prob.}\{\Omega_{k-1}=(\omega\mid\bar{\lambda})\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}\frac{(m-n_D(\lambda\mid k))!}{m!}\int_E\mathbf{C}_X^{(k)}(y,m,s_k\mid\omega,\lambda,n,\bar{Z})\mu_{s_k}^m(dy)
\end{aligned}
\tag{56}
$$

where $\bar{\lambda}$ is the unique parent of λ on $\bar{Z}$ and $(\omega\mid\bar{\lambda})\in W(\bar{\lambda},n)$ is the restriction of ω in the sense of (42). According to part [2] which has been already proven, $\mathbf{C}_X^{(k)}(\cdot\mid\omega,\lambda,n,\bar{Z})$ is invariant with respect to ω. By *Lemma 1* in the previous section, we have $\#(W(\lambda,n))=\dfrac{n!}{(n-\#(\lambda))!}$ and

$$
\text{Prob.}\{\Omega_{k-1}=\bar{\omega}\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}=\frac{(n-\#(\bar{\lambda}))!}{n!}
$$

for all $\bar{\omega}\in W(\bar{\lambda},n)$. Hence (46) and (47) follow from (56), thus completing the proof of [3].

For any $m\geq 0$, any measurable set E in $(Y_{s_k})^m$, and any $\lambda\in\mathbf{H}(Z)$, we have

$$
\begin{aligned}
&\text{Prob.}\{z(k)\in E\times\{m\},\Lambda_k=\lambda\mid\Lambda_{k-1}=\bar{\lambda},\bar{Z}\}\\
&=\sum_{n=\#(\lambda)}^{\infty}\text{Prob.}\{z(k)\in E\times\{m\},\Lambda_k=\lambda\mid\Lambda_{k-1}=\bar{\lambda},N_T=n,\bar{Z}\}\,\text{Prob.}\{N_T\mid\Lambda_{k-1}=\bar{\lambda},\bar{Z}\}
\end{aligned}
\tag{57}
$$

where $\bar{\lambda}$ is the unique parent of λ on $\bar{Z}$. Therefore, (48) and (49) follow from (56), (57) and *Lemma 1* in the previous section, completing the proof of part [4]. The reason why $\mathbf{L}_k(z(k),\lambda\mid\bar{Z},\bar{\lambda})$ is called the *likelihood of* $(z(k),\lambda)$ given $(\bar{Z},\bar{\lambda})$ is clear in (57). *Q.E.D.*

An interesting observation of the general result stated in terms of *Theorem 1* is the normalizing constant of the Bayesian formula at each level, i.e, *target system state estimation* → *number-of-target estimation* → *hypothesis evaluation*, which is always used in the next higher

level. As the final stage of this chain, we have the following corollary:

Corollary to Theorem 1: The likelihood of the current data set $z(k)$ given the past cumulative data set $Z^{(k-1)}$, or the density (in an appropriate sense) of probability distribution of the current data set $z(k)$ conditioned by the past cumulative data set $Z^{(k-1)}$, is given by

$$P_Z^{(k)}((y,m,J_k,s_k)\,|\,Z^{(k-1)}) = \sum_{\lambda\in\mathbf{H}(Z^{(k)})} \mathbf{L}_k((y,m,J_k,s_k),\lambda\,|\,\bar{\lambda},Z^{(k-1)})\,P_{\mathbf{H}}^{(k-1)}(\bar{\lambda}\,|\,Z^{(k-1)})$$

for each $(y,m)\in\bigcup_{m=0}^{\infty}(Y_{s_k})^m\times\{m\}$ and each hypothesis $\lambda\in\mathbf{H}(Z^{(k)})$ whose unique parent on $Z^{(k-1)}$ is $\bar{\lambda}$.

The proof to the above corollary is obvious from (48) and the proof of part [4] of *Theorem 1*. *Theorem 1* gives us a general multitarget tracking algorithm in which the set of variables to propagate forward is (1) all the hypotheses in $\mathbf{H}(Z)$ and all the tracks in $\mathbf{T}(Z)$, (2) evaluation $P_{\mathbf{H}}^{(k)}$ of each hypothesis, (3) number-of-target statistics $P_{N_T}^{(k)}$ given each hypothesis and (4) target system state statistics $P_{\mathbf{X}}^{(k)}$ given each data-to-data hypothesis, each number-of-targets hypothesis and each target-to-track hypothesis. By imposing additional assumptions, we can derive many different multitarget tracking algorithms from the general result shown in *Theorem 1*. One of the most interesting result is obtained when we impose the independent target assumptions together with Poisson assumption on the number of targets, which will be discussed in the next section.

5. I.I.D.-POISSON CASES

By i.i.d. (independent, identically distributed) - Poisson cases, we mean cases where several additional assumptions, the most importantly, the independence among targets and Poisson distribution of the number of targets are added to our basic assumptions, *Assumptions 1-9*. With such assumptions, the general result derived in the previous section can be greatly simplified. In short, in these cases, we may evaluate tracks and hypotheses separately, and accordingly, the filtering equation described by (39) and (40) in the previous section is also decomposed into a set of track-wise equations. First we will list our additional assumptions:

Assumption A1: For each positive integer n, the common component of the target system state space is a formal singleton, i.e., $\mathbf{X}_n^C=\{\theta\}$[25] and the space $\mathbf{X}_n^I$ for the individual target states is identically equal to some hybrid space $\mathbf{X}$, i.e.,

$$\mathbf{X}_n = (\mathbf{X})^n = \mathbf{X}\times\cdots\times\mathbf{X} \ (n \ times) \quad .$$

For each $n>0$, given $N_T=n$, the first component X of the stochastic process $(X(t),N_T)_{t\in[t_0,\infty)}$ is a system $(x_i)_{i=1}^n$ of n time-homogeneous[26] Markov processes which are independent to each other and have common initial distribution

$$q_0(dx) = Prob.\ \{x_i(t_0)\in dx\}$$

and state transition probability

$$f_{\Delta t}(dx\,|\bar{x}) = Prob.\ \{x_i(t+\Delta t)\in dx\,|\,x_i(t)=\bar{x}\} \quad .$$

In other words, we assume

$$Q_0^n(\prod_{i=1}^n dx_i) = \prod_{i=1}^n q_0(dx_i) \tag{58}$$

and

$$F_{\Delta t}^n(\prod_{i=1}^n dx_i\,|(\bar{x}_i)_{i=1}^n) = \prod_{i=1}^n f_{\Delta t}(dx_i\,|\bar{x}_i) \tag{59}$$

for each $n>0$.

Assumption A2: The *a priori* distribution of number N_T of targets is Poisson with mean ν_0, i.e.,

$$Prob.\ \{N_T=n\} = \exp(-\nu_0)\,\frac{\nu_0^n}{n!}$$

---

[25] $\mathbf{X}_n^C$ will be ignored henceforth.

[26] The time homogeneity assumption is being made again only to reduce the notational complexity.

for all $n \geq 0$.

Assumption A3: The detection of targets is target-wise independent and the detection of any individual target i depends only on its state x_i, i.e., for any data set k, any $n > 0$, any $(x_i)_{i=1}^n \in \mathbf{X}^n$ and any $D \in \mathbf{D}(n)$, we have

$$P_D(D \mid (x_i)_{i=1}^n, n, k) = \prod_{i=1}^{n} p_D(x_i \mid k)^{\chi(i;D)}(1 - p_D(x_i \mid k))^{(1-\chi(i;D))} \qquad (60)^{27}$$

with a common *target-wise detection probability function* $p_D(\cdot \mid k) : \mathbf{X} \rightarrow [0,1]$.

Assumption A4: The number of false alarms in each data set k is independent of target states or any other elements in the data set, and has distribution $p_{N_{FA}}(\cdot \mid k)$, i.e.,

$$P_{N_{FA}}(m \mid D, x, n, k) = p_{N_{FA}}(m \mid k) \qquad (61)$$

for any $(D, x, n) \in \bigcup_{n=1}^{\infty} \mathbf{D}(n) \times \mathbf{X}^n \times \{n\}$. Given the number of false alarms in data set k, the vector of false alarm measurement values is a system of i.i.d. random elements in Y_{s_k} with a common distribution having a density $p_{FA}(\cdot \mid k)$ with respect to μ_{s_k}.

Assumption A5: The error in a measurement originating from a target i in any data set k depends only on its state x_i and can be modeled by a common state-to-measurement transition probability having a density $p_M(\cdot \mid \cdot, k) : Y_{s_k} \times \mathbf{X} \rightarrow [0,\infty)$. Together with the second half of *Assumption A4*, we have

$$P_M((y_j)_{j=1}^m \mid a, m, D, (x_i)_{i=1}^n, n, k) = \Big(\prod_{i \in D} p_M(y_{a(i)} \mid x_i, k)\Big) \cdot \Big(\prod_{\substack{j=1 \\ j \in Im(a)}}^{m} p_{FA}(y_j \mid k)\Big) \qquad (62)$$

for each $n \geq 0$, each $(x_i)_{i=1}^n \in \mathbf{X}^n$, each $D \in \mathbf{D}(n)$, each $m \geq 0$, each $a \in \mathbf{A}(D, \{1,\ldots,m\})$ and each $(y_j)_{j=1}^m \in (Y_{s_k})^m$.

We should note that *Assumptions A1 - A5* are assumptions which are "additional" to *Assumptions 1 - 9* made in Section 2. For example, equations (58) - (62) satisfy the requirement

[27] For any set A, $\chi(\cdot; A)$ is the indicator function of A, i.e., $\chi(a; A)$ is 1 if $a \in A$ and 0 otherwise.

of the target interchangeability by *Assumptions 1* and *9*. Before stating the multitarget tracking algorithm resulting from these assumptions, we will introduce several helpful notations and preparations: As in the previous section, we denote the cumulative data set up to k by $Z^{(k)}$ and the corresponding cumulative measurement index set by $J^{(k)}$. Let the maximal cumulative data set be denoted by $\mathbf{Z}$, i.e., $\mathbf{Z}=\{z(k)\,|\,k>0\}=\bigcup_{k>0} Z^{(k)}$. For each k and each track $\tau\in\mathbf{T}(\mathbf{Z})$, we will associate a probability distribution on $\mathbf{X}$. First, for each k, we define

$$y[\tau\,|\,k]=\begin{cases} y_j(k) & \text{if } \tau\,|\,k=\{j\} \\ \theta & \text{if } \tau\,|\,k=\varnothing \end{cases} \tag{63}$$

where θ is used symbolically to represent the fact that track τ misses measurement at data set k. Thus $y[\tau\,|\,k]$ is the measurement value $y_j(k)$ in Y_{s_k} assigned to track τ if $\tau\,|\,k=\{j\}$ and is θ if no measurement is assigned to τ at data set k. Then we define *cumulative data set Z restricted by track* $\tau\in\mathbf{T}(\mathbf{Z})$ by

$$Z_{|\tau}=\{(y[\tau\,|\,k],k)\,|\,(z(k),k)\in Z\} \tag{64}$$

which is actually another representation of a sequence $(y[\tau\,|\,1],y[\tau\,|\,2],\ldots\ldots)$ of random elements in $\{\theta\}\cup(\bigcup_{s\in S} Y_s)$.

Consider a Markov process x on X which has initial state distribution q_0 and state transition probability $f_{\Delta t}$. Then, for each τ, we can consider $(y[\tau\,|\,k])_{k>0}$ or equivalently $Z_{|\tau}$ as a sequent of observation data from an incomplete observation mechanism which provides a measurement value in Y_{s_k} if the observation is successful and provides nothing (represented by θ) otherwise, for each k, according to the statistics specified by p_D and p_M. Such an observation mechanism can be modeled by $g(\cdot\,|\,\cdot,k):(Y_{s_k}\cup\{\theta\})\times\mathbf{X}\rightarrow[0,\infty)$ which is defined by

$$g(y\,|\,x,k)=\begin{cases} p_M(y\,|\,x,k)p_D(x\,|\,k) & \text{if } y\neq\theta \\ 1-p_D(x\,|\,k) & \text{if } y=\theta \end{cases} \tag{65}$$

for each k. For each cumulative data set Z and each track $\tau\in\mathbf{T}(\mathbf{Z})$, we denote the conditional distribution of the state of this Markov process at time t, conditioned by $Z_{|\tau}$ (Z restricted by τ) by $p_t(\cdot\,|\,Z_{|\tau})$ which we call *track state distribution*. Then we have a standard formulation of nonlinear filtering problem with a continuous-time Markov process and discrete-time observation mechanism defined by (65). Thus we have the following lemma:

Lemma 2: For each track $\tau \in \mathbf{T}(Z^{(k)})$ on each cumulative data set $Z^{(k)}$, we have

$$p_{t_k}(dx \mid Z^{(k)}_{\tau}) = c_k(y[\tau|k] \mid Z^{(k-1)}_{\tau})^{-1} g(y[\tau|k] \mid x, k) p_{t_k}(dx \mid Z^{(k-1)}_{\tau \cap J^{(k-1)}}) \tag{66}$$

where

$$c_k(y \mid Z^{(k-1)}_{\tau}) = \int_{\mathbf{X}} g(y \mid x, k) p_{t_k}(dx \mid Z^{(k-1)}_{\tau \cap J^{(k-1)}}) \tag{67}$$

for each $y \in Y_{s_k} \cup \{\theta\}$ and

$$p_{t_k}(dx \mid Z^{(k-1)}_{\tau \cap J^{(k-1)}}) = \begin{cases} \int_{\mathbf{X}} f_{t_k \to t_{k-1}}(dx \mid \bar{x}) p_{t_{k-1}}(dx \mid Z^{(k-1)}_{\tau \cap J^{(k-1)}}) & \text{if } k > 1 \\ \int_{\mathbf{X}} f_{t_k \to t_{k-1}}(dx \mid \bar{x}) q_0(dx) & \text{if } k = 1 \end{cases} \tag{68}$$

Lemma 2 provides a tracking algorithm for single-target no-false alarm cases. It is, however, nothing but a standard form of the nonlinear filtering equations based on sampled data and is obtained by straightforward application of Bayes rule. Hence, it is not necessary for us to provide a proof. The algorithm consists of the updating formula (66) with (67) and of the extrapolation formula (68). Using the notion of track state distribution, the main result of this section can be written in form of the following theorem:

Theorem 2: (I.I.D.-Poisson Cases) For each k, let $Z = Z^{(k)}$, $\bar{Z} = Z^{(k-1)}$ and $\bar{J} = J^{(k-1)}$. Then, for each hypothesis $\lambda \in H(\mathbf{Z})$ with the unique parent $\bar{\lambda}$ on $\bar{Z}$, we have

[i] *(Target System State Estimate) for each $n \geq \#(\lambda)$ and for each $\omega \in \mathbf{W}(\lambda, n)$,*

$$P_{\Sigma}^{(k)}(\prod_{i=1}^{n} dx_i \,|\omega,n,\lambda,Z) = \left(\prod_{\tau\in\lambda} p_{t_k}(dx_{\omega(\tau)}\,|Z_{\tau})\right)\left(\prod_{\substack{i=1\\ i\in \mathrm{Im}(\omega)}}^{n} p_{t_k}(dx_i\,|Z_{\emptyset})\right) \quad , \tag{69}$$

[ii] *(Number-of-Target Estimate) for each* $n \geq \#(\lambda)$,

$$P_{N_T}^{(k)}(n\,|\lambda,Z) = \exp(-\nu_k)\,\frac{\nu_k^{(n-\#(\lambda))}}{(n-\#(\lambda))!} \tag{70}$$

where $\nu_k = E(N_T - \#(\lambda)\,|\lambda,Z)$ *which is given recursively by*

$$\nu_k = c(\theta\,|\bar{Z}_{\emptyset})\,\nu_{k-1} \tag{71}$$

for each $k>0$ *with* $c(\theta\,|\bar{Z}_{\emptyset})$ *being defined by* (67), *and*

[iii] *(Hypothesis Evaluation)*

$$P_{\mathrm{H}}^{(k)}(\lambda\,|Z) = C_{\mathrm{H}}^{(k)}(Z)^{-1}\,P_{\mathrm{H}}^{(k-1)}(\bar{\lambda}\,|\bar{Z})\;L_k^{FA}(z(k)\,|\lambda)\;\prod_{\tau\in\lambda} L_k(y[\tau\,|k]\,|\bar{Z}_{\gamma\cap\bar{J}}) \tag{72}$$

where $C_{\mathrm{H}}^{(k)}(Z)$ *is the normalizing constant defined by*

$$C_{\mathrm{H}}^{(k)}(Z) = N_M(k)!\,\exp(\nu_{k-1}-\nu_k)\,P_Z^{(k)}(z(k)\,|\bar{Z}) \quad , \tag{73}$$

$L_k^{FA}(\cdot\,|\lambda)$ *is the false alarm likelihood function defined by*

$$\begin{aligned} &L_k^{FA}((y_j)_{j=1}^{m},m,J_k,S_k\,|\lambda) \\ &= (m-n_D(\lambda\,|k))!\,p_{N_{FA}}(m-n_D(\lambda\,|k))\prod_{j\in j_{FA}(m,\lambda\,|k)} p_{FA}(y_j\,|k) \end{aligned} \tag{74}$$

for each $((y_j)_{j=1}^{m},m)\in\bigcup_{m=0}^{\infty}(Y_{s_k})^{m}\times\{m\}$ *with*

$$j_{FA}(m,\lambda\,|k) = \{1,\ldots,m\}\setminus\left(\bigcup_{\tau\in\lambda}(\tau\,|k)\right) \tag{75}$$

being the set of indices for measurements which λ *hypothesizes to be false alarms, and* $L_k(\cdot\,|\bar{Z}_{\gamma\cap\bar{J}})$ *is the track-to-measurement likelihood function defined by*

$$L_k(y\,|\bar{Z}_{\tau\cap\bar{J}}) = \begin{cases} \nu_{k-1}c_k(y\,|\bar{Z}_{|\varnothing}) & \text{if } \tau\cap\bar{J}=\varnothing \\ c_k(y\,|\bar{Z}_{|\tau\cap\bar{J}}) & \text{otherwise} \end{cases} \tag{76}$$

for each $y\in Y_{s_k}\cup\{0\}$.

Proof: First we will prove [i] and [ii] by mathematical induction. [i] and [ii] are obviously true for $k=0$.[28] Suppose that [i] is true for $k-1$. Let $\lambda\in\mathbf{H}(Z)$ be an arbitrary hypothesis on Z, $n\geq\#(\lambda)$ and $\omega\in\mathbf{W}(\lambda,n)$. Let $D\in\mathbf{D}(n)$, $\bar{\omega}\in\mathbf{W}(\bar{\lambda},n)$ and $a\in\mathbf{A}(D,J_M(k))$ be determined uniquely by (42), (43) and (44). Then, it follows from the additional assumptions (59) - (62) and from the extrapolation equation (68) that, for any $m\geq 0$, any measurable set E in $\mathbf{X}^n$, we have

$$\begin{aligned}
&\int_E P_M((y_j)_{j=1}^m\,|a,m,(x_i)_{i=1}^n,n,k)\,P_{N_{FA}}(m-n_D(\lambda\,|k)\,|D,(x_i)_{i=1}^n,n,k) \\
&\qquad P_D(D\,|(x_i)_{i=1}^n,k)\int_{\mathbf{X}^n} F^n_{i_k-i_{k-1}}(\prod_{i=1}^n dx_i\,|(\bar{x}_i)_{i=1}^n)P^{(k-1)}_{\mathbf{X}}(\prod_{i=1}^n d\bar{x}_i\,|\bar{\omega},n,\bar{\lambda},\bar{Z}) \\
&= \int_E \Big(\prod_{i\in D} p_M(y_{a(i)}\,|x_i)\Big)\Big(\prod_{\substack{j=1\\ j\in Im(a)}}^m p_{FA}(y_j\,|k)\Big)\,p_{N_{FA}}(m-n_D(\lambda\,|k)\,|k) \\
&\qquad \Big(\prod_{i=1}^n p_D(x_i\,|k)^{\chi(i;D)}(1-p_D(x_i\,|k))^{(1-\chi(i;D))}\Big)\Big(\prod_{\tau\in\lambda} p_{i_k}(dx_{\bar{\omega}(\tau)}\,|\bar{Z}_{|\tau})\Big)\Big(\prod_{\substack{i=1\\ i\in Im(\bar{\omega})}}^n p_{i_k}(dx_i\,|\bar{Z}_{|D})\Big) \\
&= p_{N_{FA}}(m-n_D(\lambda\,|k)\,|k)\Big(\prod_{j\in I_{FA}(m,\lambda,k)} p_{FA}(y_j\,|k)\Big) \\
&\qquad \int_E \Big(\prod_{\substack{\tau\in\lambda\\ \omega(\tau)\in D}} p_M(y_{a(\omega(\tau))}\,|x_{\omega(\tau)},k)p_D(x_{\omega(\tau)}\,|k)p_{i_k}(dx_{\omega(\tau)}\,|\bar{Z}_{|\tau\cap\bar{J}})\Big)\Big(\prod_{\substack{\tau\in\lambda\\ \omega(\tau)\in D}}(1-p_D(x_{\omega(\tau)}\,|k))p_{i_k}(dx_{\omega(\tau)}\,|\bar{Z}_{|\tau\cap\bar{J}})\Big) \\
&\qquad\qquad \Big(\prod_{\substack{i=1\\ i\in Im(\omega)}}^n (1-p_D(x_i\,|k))p_{i_k}(dx_i\,|\bar{Z}_{|D})\Big)
\end{aligned} \tag{77}$$

where $n_D(\lambda\,|k)$ is the number of targets which are hypothesized (by λ) to be detected in data set $z(k)$ defined by (41). Then (69) follows from (77), *Lemma 2* and part [1] of *Theorem 1*, i.e, [i] is also true for k, concluding the proof of [i].

It also follows from (67) and (77) that the normalizing constant defined by (40) is given by

[28] For $k=0$, the left hand sides of (69), (70) and (72) are defined by (38'), (37') and (36'), respectively.

$$C_{\Sigma}^{(k)}((y_j)_{j=1}^{m}, m, s_k \mid \omega, \lambda, n, \bar{Z}) = p_{N_{FA}}(m - n_D(\lambda \mid k) \mid k) \Big(\prod_{j \in j_{FA}(m, \lambda \mid k)} p_{FA}(y_j \mid k) \Big)$$
$$\Big(\prod_{\substack{\tau \in \lambda \\ \omega(\tau) \in D}} c(y_{a(\omega(\tau))} \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big) \Big(\prod_{\substack{\tau \in \lambda \\ \omega(\tau) \notin D}} c(\theta \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big) \big(c(\theta \mid \bar{Z}_{\emptyset}) \big)^{n - \#(\lambda)} \tag{78}$$

where each $c(\cdot \mid \cdot)$ is defined by (67). Obviously, $\omega \in W(\lambda, n)$ is arbitrary in (78). To prove [ii], assume that [ii] is true for $k-1$. Then, for any $n \geq \#(\lambda)$ and any $\omega \in W(\lambda, n)$, it follows from (78),

$$C_{\Sigma}^{(k)}(y, m, s_k \mid \omega, \lambda, n, \bar{Z}) \frac{(n - \#(\bar{\lambda}))!}{(n - \#(\lambda))!} P_{N_T}^{(k)}(n \mid \bar{\lambda}, \bar{Z})$$
$$= d \; \frac{\nu_{k-1}^{(n - \#(\bar{\lambda}))}}{(n - \#(\lambda))!} \; c(\theta \mid \bar{Z}_{\emptyset})^{(n - \#(\lambda))} \tag{79}$$

for each $(y, m) \in \bigcup_{m=0}^{\infty} (Y_{s_k})^m \times \{m\}$ where

$$d = p_{N_{FA}}(m - n_D(\lambda \mid k) \mid k) \Big(\prod_{j \in j_{FA}(\lambda \mid k)} p_{FA}(y_j \mid k) \Big) \Big(\prod_{\substack{\tau \in \lambda \\ \omega(\tau) \in D}} c(y_{a(\omega(\tau))} \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big)$$
$$\Big(\prod_{\substack{\tau \in \lambda \\ \omega(\tau) \notin D}} c(\theta \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big) \exp(-\nu_{k-1}) \quad . \tag{80}$$

Then, according to part [3] of *Theorem 1*, (70) and (71) follow from (79) and (80), i.e., [ii] is true also for k, concluding the proof of [ii].

From (79) and (80), we can calculate the normalizing constant defined by (47) as

$$C_{N_T}^{(k)}(y, m, s_k \mid \lambda, \bar{Z}) = d \; exp(\nu_k) \, (\nu_{k-1})^{(\#(\lambda) - \#(\bar{\lambda}))} \tag{81}$$

where d is defined by (80). Let λ_{new} and λ_{old} be defined by given λ, given n and an arbitrary ω through (52) with $(a, \bar{\omega})$ being uniquely determined by (42) and (44). Then we have $\#(\lambda_{new}) = \#(\lambda) - \#(\bar{\lambda})$ and

$$\Big(\prod_{\tau \in \lambda} c(y[\tau \mid k] \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big) (\nu_{k-1})^{\#(\lambda) - \#(\bar{\lambda})}$$
$$= \Big(\prod_{\tau \in \lambda_{old}} c_k(y[\tau \mid k] \mid \bar{Z}_{\Psi \cap \bar{J}}) \Big) \Big(\prod_{\tau \in \lambda_{new}} \nu_{k-1} c_k(y[\tau \mid k] \mid \bar{Z}_{\emptyset}) \Big) \quad . \tag{82}$$

It follows from (74)-(76) and (80) - (82) that the likelihood $\mathbf{L}_k(z(k),\lambda\,|\,\bar{Z},\bar{\lambda})$ of $(z(k),\lambda)$ given $(\bar{Z},\bar{\lambda})$, defined by (49) becomes

$$\mathbf{L}_k(z(k),\lambda\,|\,\bar{Z},\bar{\lambda}) = \frac{\exp(\nu_k-\nu_{k-1})}{N_M(k)!}L_k^{FA}(z(k)\,|\,\lambda)\prod_{\tau\in\lambda}L_k(y[\tau\,|\,k]\,|\,\bar{Z}_{\dagger\cap\bar{J}}) \ . \tag{83}$$

Then [iii] follows from part [4] of *Theorem 1*. *Q.E.D.*

We should remember the empty track $\emptyset$ is always included in $\mathbf{T}(Z)$ and $p_t(\cdot\,|\,\bar{Z}_{\emptyset})$ is the common *a posteriori* distribution at time t of targets which are not detected in any data set included in Z. The reason why L_k defined by (76) is called track-to-measurement likelihood may become clear in the following interpretation of hypothesis evaluation equation (72): The posterior probability of any hypothesis λ is the product of

(1) *a priori* (with respect to the current data set) probability of λ, i.e., the past evaluation $P_{\mathbf{H}}^{(k-1)}(\bar{\lambda}\,|\,\bar{Z})$ of the unique parent $\bar{\lambda}$ of λ based on the past cumulative data set $\bar{Z}$,
(2) the likelihood $L_k^{FA}(z(k)\,|\,\lambda)$ of a set of measurements indexed by $j_{FA}(N_M(k),\lambda\,|\,k)$ being the set of false alarms in data set $z(k)$,
(3) the likelihood $L_k(y[\tau\,|\,k]\,|\,\bar{Z}_{\dagger\cap\bar{J}})=c_k(y[\tau\,|\,k]\,|\,\bar{Z}_{\dagger\cap\bar{J}})$ of a measurement $y[\tau\,|\,k]\neq\theta$ originating from a previously detected target whose track is $\tau\cap\bar{J}\neq\emptyset$ in $\bar{Z}$,
(4) the likelihood $L_k(\theta\,|\,\bar{Z}_{\dagger\cap\bar{J}})=c_k(\theta\,|\,\bar{Z}_{\dagger\cap\bar{J}})$ of a previously detected target whose track is $\tau\cap\bar{J}\neq\emptyset$ in $\bar{Z}$ being undetected in the current data set $z(k)$, and
(5) the likelihood $L_k(y[\tau\,|\,k]\,|\,\bar{Z}_{\emptyset})=\nu_{k-1}c_k(y[\tau\,|\,k]\,|\,\bar{Z}_{\emptyset})$ of a measurement $y[\tau\,|\,k]\neq\theta$ originating from a newly detected target $(\tau\cap\bar{J}=\emptyset)$,

divided by the normalizing constant. Likewise, we call $L_k(\theta\,|\,\bar{Z}_{\emptyset})=\nu_{k-1}c_k(\theta\,|\,\bar{Z}_{\emptyset})$ the likelihood of an undetected target remaining undetected.[29]

Remark 6: Part [i] of this theorem is valid even without *Assumption A2*, i.e., a Poisson distribution on number N_T of targets. In other words, [i] is a consequence of all the independence assumptions: targets, detection, measurement errors, etc. We should note that the target distributions are independent only when conditioned by a target-to-data hypothesis (ω,λ). As discussed in Section 3 (*Example 1*), when we mix (69), this independence will be generally lost. On the other hand, for parts [ii] and [iii], *Assumption A2* is crucial. Thanks to this assumption, the conditional probability (under hypothesis λ) on the number $N_T=\#(\lambda)$ of targets which are not

[29] By (71), this likelihood is nothing but the conditional expectation of the number of targets which are

detected in cumulative data set up to k is independent of λ or of any measurement $((y_j(k))_{j=1}^{N_M(k)}, N_M(k))_{k\le k}$. It is also Poisson and depends only on the information about which sensor was up at each particular time, i.e., $(t_k, \delta_k)_{k\le k}$. When we exclude *Assumption A2*, we must calculate the *a posteriori* distribution of number N_T of targets separately for each hypothesis λ, although such statistics depend only on the number $\#(\lambda)$ of tracks in λ.

Even with the additional assumptions introduced in this section, *Theorem 2* provides a general multitarget tracking algorithm which includes many existing algorithms as a proper subset in a sense which we will describe in the next section. Before closing this section, we will show a batch-processing algorithm as a corollary to *Theorem 2*.

Corollary to Theorem 2: For each K and each $\lambda \in \mathbf{H}(Z^{(K)})$, we have

$$P_{\mathbf{H}}^{(K)}(\lambda \,|Z^{(K)}) = (B_{\mathbf{H}}^{(K)})^{-1} \left(\prod_{k=1}^{K} L_k^{FA}(\lambda, z(k))\right) \left(\prod_{\tau\in\lambda} l_K(\tau, Z^{(K)})\right) \tag{84}$$

where $B_{\mathbf{H}}^{(K)}$ is the normalizing constant defined by

$$B_{\mathbf{H}}^{(K)} = P_Z(Z^{(K)})\, exp(\nu_0 - \nu_K) \prod_{k=1}^{K} N_M(k)! \quad , \tag{85}$$

and $l_K(\tau, Z^{(K)})$ is the track likelihood of τ at k defined by

$$l_K(\tau, Z^{(K)}) = \nu_0 \prod_{k=1}^{K} c_k(y[\tau\,|k]\,|Z^{(k-1)}_{\tau\cap}) \tag{86}$$

with each $c_k(\cdot\,|\cdot)$ being defined by (65).

This corollary can be proved by a straightforward repetitive application of *Theorem 2* although one must carefully handle the null track. In order to obtain track likelihood $l_K(\tau, Z^{(K)})$ for each track τ, we need to calculate the normalizing constant $c_k(\cdot\,|\cdot)$ of the filtering process according to (67), which means that we must *recursively* calculate the track state distribution according to the filtering equations (66) - (68). The track likelihood itself can be calculated recursively as

---

not detected in any data set up to and including k.

$$l_k(\tau,Z^{(k)}) = \begin{cases} \nu_0 & \text{if } k=0 \\ c_k(y[\tau|k]|Z^{(k-1)}_{\tau\cap J^{(k-1)}})\, l_{k-1}(\tau\cap J^{(k-1)},Z^{(k-1)}) & \textit{otherwise} \end{cases}$$

or

$$l_k(\tau,Z^{(k)}) = \begin{cases} \nu_k & \text{if } \tau\cap J^{(k)}=\varnothing \\ c_k(y[\tau|k]|Z^{(k-1)}_{\tau\cap J^{(k-1)}})\, l_{k-1}(\tau\cap J^{(k-1)},Z^{(k-1)}) & \textit{otherwise} \end{cases}$$

which follows immediately from (86). As shown in the last equation, the null track likelihood $l_k(\varnothing,Z^{(k)})$ is nothing but the expected number ν_k of undetected targets, which is common for all the data-to-data hypotheses. Using (85) and (86) we can actually calculate *a priori* probability of each hypothesis λ, i.e., *Prob.* $\{\Lambda=\lambda\}$. Its calculation, however, involves the whole evaluation processes. This is a reason why we may not use a simple-minded Bayesian expansion such as[30]

$$P(\lambda|Z) = \frac{P(Z|\lambda)}{P(Z)}P(\lambda) \quad .$$

6. RELATION TO EXISTING ALGORITHMS

In an appropriate sense, the algorithm provided by *Theorems 1* and *2* in the previous sections includes representative multitarget tracking algorithms which have been developed up until now, as a proper subset. This section will show this by describing the relation of our algorithm developed in the previous section to representative existing algorithms. Since Reid's algorithm described in [4] gave us the most significant motivation for this report, we will spend dispropotional space for it.

[30] *Cf. Remark 5* in *Section 3.*

6.1. Reid's Algorithm Besides all the assumptions we have made by now, in [4], the individual target state transition probability $f_{\Delta t}$ is given by a linear-gaussian model, $x_i(t_k)=F_k x_i(t_{k-1})+$*gaussian noise*, and the measurement error is also linear-gaussian as $y_j(k)=H_k x_i(t_k)+$*gaussian noise*, where the individual target state space $\mathbf{X}$ is a Euclidean space and matrices F and H are with compatible dimensions. Moreover, each measurement value space Y_s for each sensor s is implicitly assumed to be a compact set in a Euclidean space with *search* (or *scan*) volume $\mu_s(Y_s)$. In other words, each Y_s is the *field of view* of sensor s. By assuming uniform probability detection on the field of view for each sensor, we have

$$p_D(x\,|k) = p_D^{\max} \int_{Y_{s_k}} G(y-H_k x; R_k)\,dy \tag{84}$$

where $p_D^{\max}\in(0,1)$, $G(\cdot;\Xi)$ is the density of the zero-mean multidimensional gaussian distribution with variance matrix Ξ and R_k is the measurement noise variance matrix. Accordingly, the state-to-measurement transition may be modeled as

$$p_M(y\,|x,k) = \frac{G(y-H_k x; R_k)}{\int_{Y_{s_k}} G(\eta-H_k x; R_k)\,d\eta}\,. \tag{85}$$

Assume that the initial distribution q_0 of the individual targets is thinly spread over a large region and that the measurement error variance R_k is very small compared with the field of view for each sensor. Then we may conclude that each nonempty track has a state distribution which is well approximated by a gaussian distribution. Thus, for each data set k and each track $\tau\in\mathbf{T}(Z^{(k)})$ such that $\bar{\tau}=\tau\cap J^{(k-1)}\neq\emptyset$, we can approximate the track state distribution $p_{t_k}(\cdot\,|Z_{\bar{\tau}}^{(k-1)})$ by a gaussian distribution with mean $\hat{x}_{k|k-1}$ and variance $\Sigma_{k|k-1}$. Then, for $y\in Y_{s_k}$, the track-to-measurement likelihood defined by (76) becomes

$$\begin{aligned} L_k(y\,|Z_{\bar{\tau}}^{(k-1)}) &= p_D^{\max}\, G(y-H_k\hat{x}_{k|k-1}\,;\, H_k\Sigma_{k|k-1}H_k^T+R_k) \\ &= p_D^{\max}\,(2\pi)^{-\frac{\dim(y)}{2}} \det(H_k\Sigma_{k|k-1}H_k^T+R_k)^{-\frac{1}{2}} \exp(-\frac{1}{2}\chi^2) \end{aligned} \tag{86}$$[31]

where

$$\chi^2 = \|y-H_k\hat{x}_{k|k-1}\|^2_{(H_k\Sigma_{k|k-1}H_k^T+R_k)^{-1}} \tag{87}$$[32]

[31] H^T is the transpose of matrix or a vector H.

$$p_{FA}(y\,|k) = (\mu_{s_k}(Y_{s_k}))^{-1} \quad . \tag{89}$$

When we add a series of approximations mentioned above and the additional assumptions ((84), (85), (88), (89), etc.) to the general algorithm shown as *Theorem 2* in the previous section, we obtain exactly Reid's algorithm described in [4] if we define *false alarm density* by

$$\beta_{FA}(k) = \nu_{FA}(k)/\mu_{s_k}(Y_{s_k}) \tag{90}$$

and *density of previously unknown target that has been detected*[34] by

$$\beta_{NT}(k) = L_k(y[\tau\,|k]\,|Z_{|\varnothing}^{(k-1)}) = \nu_{k-1}\, c_k(y[\tau\,|k]\,|Z_{|\varnothing}^{(k-1)}) \tag{91}$$

where $y[\tau\,|k]$ is the value of the measurement which is hypothesized by track τ to originate from a newly detected target.

We should note that in [4] the left hand side of (91), i.e., $\beta_{NT}(k)$, is a constant for any given k while the right hand side is a function of measurement value $y[\tau\,|k]$. Thus (91) shows a crucial difference between our algorithm and Reid's in [4]. In fact, $\beta_{NT}(k)$ should be a function of k and the measurement value, since otherwise newly started tracks (hypothetical newly detected targets) may be given increasingly unjustifiably high possibilities. For example, suppose we are watching an area for a sufficiently long period of time with a reasonably high probability of detection. Then the chance that we see a target which is detected for the first time in the middle of the field of view is very low. This is so since if that happens such a target must have evaded detection many times while moving around or reaching the middle of field of view, which is very unlikely. To prevent this from happening, Reid proposed to adjust his β_{NT} as described in a paragraph in [4]:

> ....., a calculation of β_{NT}, the density of new (i.e., unknown) targets, is performed whenever a data set from a type 1 sensor is received. The density of new targets β_{NT} depends upon the number of times the area has been observed by a type 1 sensor and the possible flux of undetected targets into and out of the area.

Aside from the above description, there is no further explanation as to what "calculation" is performed to obtain appropriate β_{NT}. In earlier literature [14] by Reid, a rather heuristic method for calculating β_{NT} is described, in which the target state space is divided into many square cells and the inflow/outflow of undetected targets from cell to cell is considered. In effect, this kind of procedure corresponds to the calculation of expected number ν_k of undetected targets and the state distribution $p_t(\cdot\,|Z_{|\varnothing})$ of null track.

[34] Reid's terminology in [4].

The derivation of Reid's algorithm in [4] is based upon the following Bayesian expansion: First let

$$N_{NT}(k) = \#(\Lambda_k) - \#(\Lambda_{k-1})$$

be the number of newly detected targets in data set k. Reid then assumed that $N_{NT}(k)$ has a Poisson distribution with a given mean[35] $\nu_{NT}(k)$ and that the measurement value $y_{A_k(i)}$ originating from each newly detected target i is uniformly distributed on the field of view, Y_{s_k}. With the uniform detection assumption, the density of newly detected target is calculated as

$$\beta_{NT}(k) = \nu_{NT}(k) / \mu_{s_k}(Y_{s_k}) \quad .$$

Let

$$N_{DT}^{old}(k) = \# \left(\Omega_{k-1}(\Lambda_{k-1}) \cap I_{DT}(k) \right)$$

be the number of targets which are detected in some data set included in $Z^{(k-1)}$ and also in the *current* data set $z(k)$. Then, again with the uniform detection assumption, we have

$$Prob.\{N_{DT}^{old}(k)=n \mid \Lambda_{k-1}=\bar{\lambda}, Z^{(k-1)}\} = \frac{\#(\bar{\lambda})!}{n!\,(\#(\bar{\lambda})-n)!} (p_D^{\max})^n (1-p_D^{\max})^{(\#(\bar{\lambda})-n)} \tag{92}$$

for each $\bar{\lambda} \in \mathbf{H}(Z^{(k-1)})$, each integer n such that $0 \le n \le \#(\bar{\lambda})$. In other words, the probability distribution of $N_{NT}^{old}(k)$ conditioned by Λ_{k-1} becomes a binomial distribution with mean $\#(\Lambda_{k-1}) p_D^{\max}$. Then, conditioned by $N_{NT}(k)$ and $N_{DT}^{old}(k)$, the probability distribution of the number $N_M(k)$ of measurements in the current data set $z(k)$ is determined by the number-of-false-alarm probability function $p_{N_{FA}}$, i.e., (88). Assuming that $N_{DT}^{old}(k)$, $N_{NT}(k)$ and $N_{FA}(k)$ are independent when conditioned by $(\Lambda_{k-1}, Z^{(k-1)})$, we have

$$\begin{aligned} Prob.\{N_{DT}^{old}(k)=n_{DT}, N_{NT}(k)=n_{NT}, N_{FA}(k)=n_{FA} \mid \Lambda_{k-1}=\bar{\lambda}, Z^{(k-1)}\} &= e^{-(\nu_{FA}(k)+\nu_{NT}(k))} \\ \frac{(\nu_{FA}(k))^{n_{FA}}}{n_{FA}!} \frac{(\nu_{NT}(k))^{n_{NT}}}{n_{NT}!} \frac{\#(\bar{\lambda})!}{n_{DT}!\,(\#(\bar{\lambda})-n_{DT})!} &(p_D^{\max})^{n_{DT}} (1-p_D^{\max})^{\#(\bar{\lambda})-n_{DT}} \end{aligned} \tag{93}$$

To obtain the final form of Reid's algorithm, the following two equations are used:

[35] $p_D^{\max}$ is included in $\nu_{NT}(k)$ since $N_{NT}(k)$ is the number of newly targets. Roughly speaking, we have $\nu_{NT}(k) = p_D^{\max} \nu_{k-1}$.

$$\begin{aligned} &Prob.\{\Lambda_k=\lambda \mid N_{FA}(k)=n_{FA}, N_{DT}^{old}=n_{DT}, N_{NT}=n_{NT}, \Lambda_{k-1}=\bar{\lambda}, Z^{(k-1)}\} \\ &= \frac{n_{FA}!\, n_{NT}!}{(n_{FA}+n_{DT}+n_{NT})!} \frac{n_{DT}!\,(\#(\bar{\lambda})-n_{DT})!}{\#(\bar{\lambda})!} \end{aligned} \tag{94}$$

and

$$\begin{aligned} &Prob.\{(y_j(k))_{j=1}^m \in (dy_j)_{j=1}^m \mid N_M(k)=m, \Lambda_k=\lambda, Z^{(k-1)}\} \\ &= (1/\mu_{s_k}(Y_{s_k}))^{n_{FA}(m\lambda\mid k)+n_{NT}(\lambda\mid k)} \\ &\quad \Big(\prod_{\substack{\tau\in\lambda_{old} \\ \tau\mid k=(j)}} G(y_j - H_k \hat{x}_{t_k}(Z^{(k-1)}_{\tau\cap J^{(k-1)}}) ; H_k \Sigma_{t_k}(Z^{(k-1)}_{\tau\cap J^{(k-1)}}) H_k^T + R_k) \Big) \prod_{j=1}^m \mu_{s_k}(dy_j) \end{aligned} \tag{95}$$

where $\hat{x}_{t_k}(Z^{(k-1)}_{\tau\cap J^{(k-1)}})$ and $\Sigma_{t_k}(Z^{(k-1)}_{\tau\cap J^{(k-1)}})$ are the mean and the variance of the track state distribution $p_{t_k}(\cdot \mid Z^{(k-1)}_{\tau\cap J^{(k-1)}})$ which is assumed to be gaussian.

Reid obtained his final form by multiplying the three key equations, (93) - (95). In all of these equations, target interchangeability is used implicitly. With closer observation, we may question the exactness of these equations: For example, when we integrate (95) over $(Y_{s_k})^m$, the integral will be strictly less than one while it should be exactly one. This is so because Y_{s_k} is a compact and hence a proper subset of a Euclidean space. Moreover, (93) and (94) are valid only when the probability of detecting targets which have been detected before is the same for all such targets. For example, if a target leaves the field of view of a sensor, such an assumption certainly invalid. The same conclusion may be said when a target is detected by a sensor but is outside of the field of view of another sensor, or when a target is on the edge of the field of view of a sensor. On the other hand, according to our formulation shown in the previous sections, the statistics $N_{DT}^{old}(k)$ can be calculated as

$$\begin{aligned} &Prob.\{N_{DT}^{old}=n \mid \Lambda_{k-1}=\bar{\lambda}, Z^{(k-1)}\} \\ &= \sum \Big\{ \prod_{\tau\in\bar{\lambda}} \Big(\int_{\mathbf{X}} p_D(x\mid k) p_{t_k}(dx \mid Z^{(k-1)}_{\bar{\tau}}) \Big)^{\in(\bar{\tau})} \Big(1 - \int_{\mathbf{X}} p_D(x\mid k) p_{t_k}(dx \mid Z^{(k-1)}_{\bar{\tau}}) \Big)^{1-\in(\bar{\tau})} \\ &\qquad \Big| \in : \bar{\lambda} \to \{0,1\}, \sum_{\bar{\tau}\in\bar{\lambda}} \in(\bar{\tau}) = n \Big\} \end{aligned} \tag{92'}$$

Similarly, the statistics on number $N_{NT}(k)$ of newly detected targets on data set k is calculated as

$$Prob.\{N_{NT}(k)=n_{NT} \mid \Lambda_{k-1}=\bar{\lambda}, Z^{(k-1)}\} = \exp(-\nu_{k-1}p)\,\frac{(\nu_{k-1}p)^{n_{NT}}}{n_{NT}!} \tag{96}$$

for any $\bar{\lambda}\in \mathbf{H}(Z^{(k-1)})$ where

$$p = \int_{\mathbf{X}} p_D(x \mid k)\, p_{t_k}(dx \mid Z_{\emptyset}^{(k-1)}) \quad .$$

Namely, in our formulation, the Poisson assumption on the total number of targets plus independence assumptions together imply that the distribution of the number $N_{NT}(k)$ of *newly detected targets* at k is Poisson. While, in [4], (96) is used as an assumption. Traditionally, newly detected targets are treated by a very ambiguous notion of "Poisson arrival" with which any underlying target model is not clearly stated. Our formalism is based upon a very simple observation that, if a target is newly detected, it must exist. This includes the case where a target is born in the middle of the field of view of a sensor. In other words, our concept of existence is independent of time.

Reid also extended his linear-gaussian result outlined above to non-gaussian cases (totally discrete space and hybrid space) in [12], [15] and [16]. Besides the problem mentioned above, the uniform detection assumption may be inadequate in some applications. For example, when a surveillance region includes masking areas or sensor's detection is highly dependent on a target state component (such as MTI-type radars, emission-detection type sensors, radars with cross-section sensitive detection, etc.), this assumption may be unacceptable. In such a case, the detection probability function must be included in an integrand of the track-to-measurement likelihood rather than a constant multiplier, and accordingly, the track state distribution must be updated by assumption that the hypothesized target has evaded the detection. When we ignore this consideration, however, there is no need for re-evaluation of either hypotheses or track state distributions by a no-measurement data set $(\theta,0,t,s)$. This is why Reid separately treats type-2 sensors which generates only data sets each of which contains at most one measurement. Apparently, when no-measurement data sets are valuable information, such treatment is inadequate.

6.2. PDA and JPDA As shown in the previous subsection, the most difficult part of multi-target tracking problems arises when the number of targets is not known so that we must hypothesized each measurement as a potential new target which has not been recognized before. Thus if the number of target is known, considerable simplification is possible. The PDA (*Probabilistic Data Association* [17]) and the JPDA (*Joint Probabilistic Data Association* [18] and [19]) algorithms are based upon the assumption of known number of targets. Moreover, one of the underlying assumptions for these algorithms may be characterized as *targets with a priori identification* as discussed later. First, we will show that our algorithm can be reduced to PDA or

JPDA algorithm with additional assumptions:

To do this, aside from all the assumptions made in Sections 4 and 5, we add an extra sensor which we may call "super sensor" providing *a priori* information on targets. Let $\bar{S}$ be $S \cup \{\bar{s}\}$ and $\bar{s} \notin S$. Assume that, at $k=0$ and only at $k=0$, sensor $\bar{s}$ provides a data set $((y_i(0))_{i=1}^{n}, n, \bar{s})$ and that, in this special data set, the probability of detecting all the targets is *one* and there is no false alarm. The effect of having this special data set is twofold; (1) The number of targets is exactly known, i.e., $N_T = n$, and (2) targets are labeled by measurement index $(i,0)$. Suppose, as a result of update-by-measurement, we have the track state distribution $p_{t_0}(\cdot \mid Z^{(0)}_{\{(i,0)\}})$ for track $\{(i,0)\}$. For each data set $k>0$, because *Prob.* $\{N_T = n\}=1$ the likelihood of any measurement originating from a newly detected target is always zero. Therefore, when we extend every cumulative data set to include the super sensor data set $k=0$, all the hypotheses in $\mathbf{H}(Z)$ for any cumulative data set Z with positive *a posteriori* probability have the same number of tracks which uniquely correspond to each other through their unique predecessors on $Z^{(0)}$. On the other hand, $\mathbf{H}(Z)$ is a mutually distinct and collectively exhaustive set of hypotheses on Z, or $\{\{\Lambda_k = \lambda\} \mid \lambda \in \mathbf{H}(Z)\}$ is a partition of the underlying probability space which is measurable with respect to the σ-algebra generated by cumulative data set Z. Therefore, we can *combine*[36] data-to-data hypotheses provided we know which track corresponds to which track in each hypothesis. For example, when $\lambda_i = \{\tau_1^i, \tau_2^i\}$ $(i=1,2)$ represents a two hypotheses on a cumulative data set $Z^{(k)}$, if we know τ_j^1 and τ_j^2 originate from the same target for each j, then we can combine these two hypotheses into one hypothesis $\lambda = \{\tau_1, \tau_2\}$ so that each "combined" track τ_j has track state distribution

$$p_{t_k}(\cdot \mid Z^{(k)}_{\tau_j}) = p_{t_k}(\cdot \mid Z^{(k)}_{\tau_j})\, P_{\mathbf{H}}(\lambda_1 \mid Z^{(k)}) + p_{t_k}(\cdot \mid Z^{(k)}_{\tau_j})\, P_{\mathbf{H}}(\lambda_2 \mid Z^{(k)}) \quad . \tag{97}$$

Having included a "super" sensor, we know exactly which track goes to which track. Therefore, we can actually combine all the hypotheses at each k thereby propagating only one hypothesis to the next data set. Moreover, by manipulating the individual target state space, we can have different kinds of target dynamics assuming that such information is also provided by the "super" sensor. From another view point, by having the "super" sensor, the targets are given *a priori* identification. Therefore, for "non-super" sensors, the origin of every measurement is considered in reference to such target identification. This kind of approach may be appropriate when the targets of interest are positively identifiable and their number is small. We have thus derived an extended version of PDA ($N_T=1$) and JPDA ($N_T>0$) from our general algorithm shown in Section 6. We also have shown that, in an appropriate sense, multitarget tracking without *a priori* identification (labeling) is more general than that with *a priori* identification.

Like Reid's algorithm, the PDA and JPDA algorithms assume the linear gaussian target

[36] Combining hypotheses is one of the critical topics in implementing any multi-hypothesis multitarget tracking algorithm and will be treated more thoroughly in Part II of this report.

dynamics and measurement mechanisms. PDA uses so-called "improper" distribution[37] and JPDA uses the Poisson number-of-false-alarm assumption (88). Also each track state distribution is assumed to be gaussian. When combining hypotheses, track state distribution of each "combined" track is therefore the sum-of-gaussian distribution with *a posteriori* probabilities of hypotheses as weights, as in (97). As in Reid's algorithm, such a sum-of-gaussian distribution is then approximated by a gaussian distribution both in PDA and JPDA. Therefore, the JPDA algorithm may be obtained also from Reid's algorithm by letting the newly detected target density be zero and by assuming a "super sensor" as discussed above, when hypothesis combining is performed to its extreme degree (combine-all). However, we must admit that the above discussions are slightly artificial and lack in theoretical rigor by introducing an artificial "super sensor." More rigorous treatment of targets with *a priori* identification (labeling) should be done by an analysis which is similar to that in Sections 3 and 4 but without *a priori* interchangeability assumption, *Assumption 2*. In such an analysis, we must evaluate target-to-data hypotheses rather than data-to-data hypotheses. Then hypothesis combining is performed on target-to-data hypotheses. In effect, this combination is equivalent to the mixture of target system state distribution via target-to-track hypotheses, which was discussed in Section 3 in terms of *Example 1*. Therefore, the JPDA algorithm involves the two major approximations: (1) approximation of sum-of-gaussian distributions by gaussian distributions and (2) approximation of dependent distributions (resulting from target-to-data hypothesis mixture – or hypothesis combining) by independent distributions. To the best knowledge of the authors, the second kind of approximation involved in the JPDA algorithm has not been noticed up to date.

Several minor differences of the JPDA algorithm from Reid's algorithm may also be observed. One of the differences is the treatment of validation regions. In Reid's algorithm, the probability of a measurement falling out of the validation region of any track is ignored. As mentioned in Section 6.1, the data validation can be appropriately viewed as an approximation in Reid'd algorithm. On the other hand, in PDA and JPDA algorithms, the probability of "real" measurement falling out the validation region is explicitly considered. Moreover, a delicate conceptual difficulty arises from the fact that, in [17] - [19], hypotheses are formed based upon the result of data validation. This results in a slight difference between the PDA and JPDA algorithms, i.e., when we let the number of target be one, the JPDA algorithm does not coincide with the PDA algorithm[38]. Nonetheless, unless validation regions are inadequately small[39], the treatment of data validation should not affect performance of any multitarget tracking algorithm and such delicate difference resulting from different treatments of data validation should

[37] An improper distribution is a probability distribution P on a measure space $(X,\mathcal{B},\mu)$ having a constant density p on X with respect to μ when $\mu(X)=\infty$. However, apparently such a probability distribution never exists. More plausible modeling will be obtained when we assume that, for each data set k, the number-of-false-alarm probability function $p_{N_{FA}}(\cdot\,|k)$ is constant on $\{0,1,.....,N_{FA}^{\max}(k)\}$ with *a priori* upper bound $N_{FA}^{\max}(k)$ When there is no newly detected target in the data set, this upper bound does not affect the hypothesis evaluation because of the cancellation as far as the number of measurement $N_M(k)$ does not exceed $N_{FA}^{\max}(k)$. Therefore, we can chose any upper bound even after the data set is received.

[38] After adjusting both algorithms due to the difference of the number-of-false-alarm distribution assumptions.

[39] In such a case, we can easily imagine that any algorithm may not perform well.

be negligible. In the JPDA algorithm, the probability of detection is assumed to be a function of "targets." However, for each given target, the detection probability is a constant and independent of its state like in Reid's algorithm. As recognized widely, both PDA and JPDA algorithms implicitly assume separate a "track initiator" which provides all the *a priori* information, i.e., the number of targets and state estimate for each target. Then we may call algorithms such as PDA and JPDA "track continuation." It is obvious, however, these two processes, track initiation and track continuation, cannot be carried out independently in general cases. Let us call a system for track continuation a *track extender* tentatively. For example, a newly detected target may produce a measurement in a validation region of a track maintained by a track extender and may be ignored by both because a track extender considers it as a false alarm and a track initiator considers it as a continued track. In such a case, both must negotiate. In [4], Reid shows that his algorithm can be used for track initiation and continuation simultaneously. The clustering procedure described in [4] and discussed in Part II of this report then provides a way to perform these two processes separately as far as it is possible. As shown in Part II of this report, however, the Poisson assumption of the number of false alarms in each data set is a crucial assumption to enable the clustering procedure.

6.3. Morefield's Algorithm Morefield's paper [11] motivated us in selecting appropriate definitions of tracks and hypotheses. As mentioned before, our definitions are different from his in that we have separated the measurement-value information from the number-of-measurement information. By this separation, we have succeeded in treating tracks and hypotheses mathematically rigorously as shown in Sections 4 and 5. Although this report was motivated considerably by Reid's paper [4], it does not clearly defined tracks and hypotheses. In [17] - [19] for the PDA and JPDA algorithms, only target-to-data hypotheses are in effect considered. Before discussing Morefield's algorithm which uses a 0-1 linear programming technique, we will show a general batch-processing algorithm which is obtained by adding the Poisson number-of-false-alarm assumption (88) for each data set.

With this additional assumption, for each k, false alarm likelihood L_k^{FA} defined by (73) in Section 5 becomes

$$L_k^{FA}(z(k)\,|\,\lambda) = \exp(-\nu_{FA}(k)) \prod_{j \in j_{FA}(N_M(k),\lambda\,|\,k)} \nu_{FA}(k) p_{FA}(y_j(k)\,|\,k) \tag{98}$$

for each hypothesis $\lambda \in \mathbf{H}(Z^{(k)})$ where $j_{FA}(\cdot,\cdot\,|\,k)$ is defined by (75) in Section 5 and $\nu_{FA}(k)$ is the expected number of false alarms in data set k. For each data set K and each track $\tau \in \mathbf{T}(Z^{(K)})$, define the *modified track likelihood* by

$$\tilde{\lambda}^{(\tau, Z^{(K)})} = \nu_0 \prod_{k=1}^{K} \frac{c_k(y[\tau\,|\,k]\,|\,Z^{(k-1)}_{|\tau \cap j(k-1)})}{\left(\nu_{FA}(k) p_{FA}(y[\tau\,|\,k]\,|\,k)\right)^{\in(y[\tau\,|\,k])}} \tag{99}$$

where

$$\in(y) = \begin{cases} 1 & \text{if } y \neq \theta \\ 0 & \text{if } y = \theta \end{cases} \tag{100}$$

Then it is easily seen that we can restate the result of *Corollary to Theorem 2* in Section 5 as

$$P_{\mathrm{H}}^{(K)}(\lambda \mid Z^{(K)}) = (\tilde{\Phi}^{(K)})^{-1} \prod_{\tau \in \lambda} \tilde{J}(\tau, Z^{(K)}) \tag{101}$$

with

$$\tilde{\Phi}^{(K)} = P_Z(Z^{(K)}) \exp(\nu_0 - \nu_K) \exp(\sum_{k=1}^{K} \nu_{FA}(k)) \prod_{k=1}^{K} N_M(k)! \tag{102}$$

for each K and each $\lambda \in \mathrm{H}(Z^{(K)})$. In (84) and (101), the product over the empty set is defined as one as usual. Take the logarithm of (101) and ignore the normalizing constant. Then we have a function $h_K : \mathrm{H}(Z^{(K)}) \rightarrow \mathbb{R}$ defined by

$$h_K(\lambda) = \sum_{\tau \in \lambda} \log \tilde{J}(\tau, Z^{(K)})) = \sum_{\tau \in \mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}} \log \tilde{J}(\tau, Z^{(K)})) \, \chi(\tau ; \lambda) \tag{103}$$

for every $\lambda \in \mathrm{H}(Z^{(K)})$.

Thus the problem of obtaining the maximum *a posteriori* probability (MAP) hypothesis λ at K is equivalent to that of maximizing $h_K(\lambda)$. In (103),

$$(\chi(\tau ; \lambda))_{\tau \in \mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}} \in \{0,1\}^{\mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}}$$

for each $\lambda \in \mathrm{H}(Z^{(K)})$ or each λ can be considered as a 0-1 vector with dimension $\#(\mathrm{T}(Z^{(K)}) \setminus \{\varnothing\})$. Namely, $\mathrm{H}(Z^{(K)})$ is isomorphic to a subset

$$\mathbf{F} = \left\{ \mathbf{x} \in \{0,1\}^{\mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}} \;\middle|\; \mathbf{x}_{\tau_1} + \mathbf{x}_{\tau_2} \leq 1 \text{ for } \textit{all } (\tau_1, \tau_2) \textit{ such that } \tau_1 \neq \tau_2 \text{ and } \tau_1 \cap \tau_2 \neq \mathcal{Z} \right\}$$

of $\{0,1\}^{\mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}}$. Define a binary matrix $\mathbf{A} \in \{0,1\}^{(\mathrm{T}(Z^{(K)}) \setminus \{\varnothing\}) \times J^{(K)}}$ by

$$\mathbf{A}_{(\tau,(j,k))} = \lambda((j,k);\tau)$$

for all $\tau \in \mathbf{T}(Z^{(K)}) \setminus \{\emptyset\}$ and all $(j,k) \in J^{(K)}$. Then we have

$$\mathbf{F} = \{\mathbf{x} \in \{0,1\}^{\mathbf{T}(Z^{(K)}) \setminus \{\emptyset\}} \mid \mathbf{A}\mathbf{x} \leq 1\}$$

where $1 \in \{0,1\}^{J^{(K)}}$ is the all-one binary vector. Thus the problem of maximizing (101) or equivalently (103) over the set of all the hypotheses in $\mathbf{H}(Z^{(K)})$ can be rewritten as a 0-1 linear programming

$$\begin{aligned} &\textit{find} \quad \mathbf{x} \in \{0,1\}^{\mathbf{T}(Z^{(K)}) \setminus \{\emptyset\}} \quad \textit{which} \\ &\textit{maximizes} \quad \mathbf{c}^T \mathbf{x} \\ &\textit{subject to} \quad \mathbf{A}\mathbf{x} \leq 1 \end{aligned}$$

where $\mathbf{c} \in \mathbb{R}^{\mathbf{T}(Z^{(K)}) \setminus \{\emptyset\}}$ is defined by

$$\mathbf{c}_\tau = \log \tilde{l}(\tau, Z^{(K)})$$

for all $\tau \in \mathbf{T}(Z^{(K)})$. To solve the above linear programming, we need a special algorithm which is extensively discussed in [11]. Although we are not certain, some of algorithms for solving classical assignment problems such as Munkres' algorithm[20] or modified Munkres'[21] may be applicable or modified to solve the above problem formulated as a linear programming.

The modified track likelihood can be recursively calculated as

$$\tilde{l}(\tau, Z^{(k)}) = \begin{cases} \nu_0 & \text{if } k=0 \\ \dfrac{c_k(y[\tau|k] \mid Z^{(k)}_{\tau \cap J^{(k)}})}{\nu_{FA}(k) p_{FA}(y[\tau|k]|k)} & \textit{else if } y[\tau|k] \neq \emptyset \\ c_k(\theta \mid Z^{(k)}_{\tau \cap J^{(k)}}) & \textit{otherwise} \end{cases}$$

or

$$\tilde{l}(\tau, Z^{(k)}) = \begin{cases} \nu_k & \text{if } \tau \cap J^{(k)} = \varnothing \\ \dfrac{c_k(y[\tau|k] \,|\, Z^{(k)}_{\tau \cap J^{(k)}})}{\nu_{FA}(k) p_{FA}(y[\tau|k]|k)} & \text{else if } y[\tau|k] \neq \varnothing \\ c_k(\theta \,|\, Z^{(k)}_{\tau \cap J^{(k)}}) & \text{otherwise} \end{cases}$$

for each k and each $\tau \in \mathbf{T}(Z^{(k)})$. Thus (101) also provides algorithms in which track likelihood for each track is recursively calculated and hypothesis evaluation is performed by multiplication of track likelihoods. In [11], like Reid's (in [4]), PDA, JPDA and other many algorithms, target dynamics and measurement mechanisms are modeled as linear-gaussian. Therefore, we can repeat the process shown in Section 6.1 to reduce the general algorithm derived in Section 5 to Morefield's algorithm. However, two comments may be noteworthy: (1) The probability of detection is totally ignored in the track likelihood in [11]. (2) The target density information is also ignored in [11]. When the probability of detection is reasonably high and the value of track likelihood is dominated by gaussian terms, the above (1) may not be any deteriorating factor. The above (2) was given some consideration in [11] as part of *a priori* information represented by $P(\lambda)$ which we cannot define and/or calculated easily as mentioned in the last paragraph in Section 5. We should note that our formulation in terms of (101) does include all the *a priori* information regarding targets and sensors.

6.4. Goodman's General Model In [5] and [8], Goodman describes a model which is a generalization of conventional linear-gaussian models and derived an algorithm based on it. The final results in both papers are very complicated and are spread out through over more than ten equations. We will try to extract the essence of his approaches and results as much as possible. Since the result in [8] is incorporated into [5], we will discuss only [5]. Our assumptions, *Assumptions 1 - 9* and *Assumptions A1 - A5*, made in Sections 2 and 5 are all shared by the set assumptions made in [5] except for those listing below[40]:

(1) The number of targets is possibly infinity.
(2) The detection of an individual target is represented by a Markov process.
(3) Possibly unresolved measurements, i.e., merged measurements are considered.

One of the causes which make [5] and [8] very complicated and difficult to follow is the inclusion of the targets' birth-death processes. This may be handled more simply by setting the individual target state space as

$$\mathbf{X} = \mathbf{X}_{geolocational} \times \{dead, alive\} \times \mathbf{X}_{other\ discrete\ components} .$$

Targets are indexed by positive integers and every positive integer corresponds to a target which

[40] Plus other minor details which the authors may have overlooked.

may "emerge." Therefore, (9) in Section 2 may be modified as

$$I_T^{\infty} = \{i \in \mathbb{Z}_+ \mid \textit{target } i \textit{ once lives}\} \tag{9'}$$

and $N_T^{\infty} = \#(I_T^{\infty})$. Then, in [5], the probability of $N_{DT}^{\infty} = \infty$, i.e., $I_T^{\infty} = \mathbb{Z}_+$, may not be zero. This may cause some mathematical concern because $\mathbb{R}^{\infty}$ is not a locally compact space with the usual direct-product topology. However, with an appropriate assumptions on the birth-death process, at any finite time t, the number of targets which has been born up until t (including already dead ones) is always finite. Also the inclusion of possibility of merged measurements makes the discussions in [5] very complicated. We will therefore retain our no-merged-measurement assumption in the following discussion. The above assumption (2) in [5] is rather questionable. In Assumption (I-6) of in Section 11.4 of [5], it is stated that the detection probability of a live target may be function of its state and "typically could be a monotonically decreasing function of the distance between the positional components and the location of the centroid, for example, of sensor system." This statement may conflict with the markovian assumption of target detection. The authors believe that the conditional independence assumptions, i.e., *Assumption 4* in Section 2 and *Assumption A3* are more plausible and realistic. The target dynamics and measurement mechanisms in [5] are modeled by "so-called" event-driven linear systems, i.e., for target i, we have

$$dx_i^c(t) = F(x_i^d(t), t)x_i^c(t)dt + G(x_i^d(t), t)dw_i(t) + u_i(x_i^d(t), t)dt$$

and

$$y_i^c(t_k) = H(x_i^d(t_k), t_k)x_i^c(t_k) + v_i(t_k)$$

where $x_i^c(t)$ and $x_i^d(t)$ are the continuous (geolocational) and the discrete[41] parts of the i-th target, respectively. $y_i^c(t_k)$ *is the continuous part of the measurement in data set* k originating from target i while $y_i^d(t_k)$ have a certain transition probability from $(x_i^{(c)}(t_k), x_i^d(t_k))$.

The uncertainty of measurement origins and hypotheses on it is modeled as follows: For each data set k, let the set of measurement values originating from detected targets and false alarms be denoted by

$$((y_i^{FA}(k))_{i=1}^{N_{FA}(k)}, N_{FA}(k))$$

and

[41] {*dead, alive*} component being excluded.

$$((y_i^{DT}(k))_{i \in I_{DT}(k)}, I_{DT}(k))$$

where $N_{FA}(k)$ is the random integer representing the number of false alarms and $I_{DT}(k)$ is the random set of detected-target indices. Then, for each k, the uncertainty of measurement origins is modeled by a random function α_k which has formal disjoint union

$$\left(\{1,.....,N_{FA}(k)\}\times\{FA\}\right) \cup \left(I_{DT}(k)\times\{DT\}\right)$$

as its domain and takes values in

$$J_M(k) = \{1,....,N_M(k)\}$$

where $N_M(k)=N_{DT}(k)+N_{FA}(k)=\#(I_{DT}(k))+\#(J_{FA}(k))$ is the number of measurement in data set k. Every α_k is then assumed to be one-to-one. Conditioned by its domain and range, every realization of α_k is assumed to be equally probable. For each data set K, define a random set $\mathbf{I}^{(K)}$ by

$$\mathbf{I}^{(K)} = \left(\bigcup_{k=1}^{K} I_{DT}(k)\times\{DT\}\right) \cup \{0\} \quad .$$

Let $\mathbf{J}^{(K)}$ be the power set of the cumulative measurement index set $J^{(K)}$, i.e.,

$$\mathbf{J}^{(K)}=\{\mathrm{J} \mid \mathrm{J}\subseteq J^{(K)}\} \quad .$$

Define a random mapping $Q^{(K)}$, from $\mathbf{I}^{(K)}$ into $\mathbf{J}^{(K)}$, by

$$Q^{(K)}(\mathrm{i}) = \begin{cases} \{(\alpha_k(\mathrm{i}),k) \mid 1\le k\le K\} & \text{if } \mathrm{i} \in \bigcup_{k=1}^{K} I_{DT}(k)\times\{DT\} \\ \bigcup_{k=1}^{K}\{(\alpha_k(i,FA),k) \mid 1\le i\le N_{FA}(k)\} & \text{if } \mathrm{i}=0 \end{cases}$$

whose realizations are called *labeled partitions*. Then we have

$$Q^{(K)}(\mathrm{i}) \neq \varnothing \quad \text{for } \textit{all } \mathrm{i}\in\mathbf{I}^{(K)}\setminus\{0\} \quad ,$$

$$Q^{(K)}(\mathrm{i}) \bigcap Q^{(K)}(\mathrm{i}') = \varnothing \quad \text{for } \textit{all } (\mathrm{i},\mathrm{i}') \in \mathrm{I}^{(K)} \times \mathrm{I}^{(K)} \textit{ such that } \mathrm{i} \neq \mathrm{i}' \text{ and}$$

$$\bigcup_{\mathrm{i} \in \mathrm{I}^{(K)}} Q^{(K)}(\mathrm{i}) = J^{(K)} \quad .$$

Namely, the image of random function $Q^{(K)}$ is a partition[42] of $J^{(K)}$ but, as a function, $Q^{(K)}$ attaches labels to each element of the partition. Each label is either a target index ($\mathrm{i} > 0$) or a common false alarm label. Since the false alarms are independent, this formulation is equivalent to (Ω_K, Λ_k) defined by Section 3. In other words, a labeled partition in [5] is another representation of a target-to-data hypothesis defined in Section 3.

Then the evaluation of each labeled partition, or equivalently of each target-to-data hypothesis, done in [5] generally follows the steps taken in the derivation of Reid's algorithm in [4] as

$$P(z(k), Q^{(k)} \mid Z^{(k-1)}, Q^{(k-1)}) = P(z(k) \mid N_M(k), Q^{(k)}, Z^{(k-1)}) P(Q^{(k)}, N_M(k) \mid Q^{(k-1)}, Z^{(k-1)}) \quad .$$

The second factor on the right hand side of the above equation is calculated by the measurement error and false alarm value models while the second factor is subsequently expanded by number $N_T(k) = \#(I_T^{(k)}) = \#(\bigcup_{k=1}^{k} I_T(k))$ of targets born up to t_k and number $N_{new}(k)$ of newly born targets as

$$P(Q^{(k)}, N_M(k) \mid Q^{(k-1)}, Z^{(k-1)}) = \sum_{N_T(k)} \sum_{N_{new}(k)} P(Q^{(k)}, N_M(k), N_T(k), N_{new}(k) \mid Q^{(k-1)}, Z^{(k-1)})$$

Each term of the right hand side is then expanded as

$$\begin{aligned} & P(Q^{(k)}, N_M(k), N_T(k), N_{new}(k) \mid Q^{(k-1)}, Z^{(k-1)}) \\ & \quad = P(Q^{(k)} \mid N_{NT}(k), N_{DT}^{old}(k), N_{FA}(k), N_T(k), N_{new}(k)) \\ & \quad\quad P(N_{NT}(k), N_{DT}^{old}(k), N_{FA}(k) \mid N_T(k), N_{new}(k)) \\ & \quad\quad P(N_T(k) \mid N_{new}(k), Q^{(k-1)})\, P(N_{new}(k) \mid Q^{(k-1)}) \end{aligned}$$

Namely, like in the derivation of Reid's algorithm discussed in Section 6.1, a crucial step to derive the multitarget tracking algorithm in [5] is a Bayesian expansion by means of the number

[42] When the set of false alarms, $Q^{(K)}(0)$, is not empty. If it is empty, $\{Q^{(k)}(\mathrm{i}) \mid \mathrm{i} \in \mathrm{I}^{(K)} \setminus \{0\}\}$ is a partition of $J^{(K)}$.

$N_{NT}(k)$ of newly detected targets, the number $N_{DT}^{old}(k)$ of targets which has been detected before and detected again at k and the number $N_{FA}(k)$ of false alarms. As mentioned in Section 6.1, this kind of expansion is useful only when the detection probability is not functionally related to target states. Thus the comments made in Section 6.1 may be relevant also to [5].

As oppose to Reid [4] and Morefield [11] which evaluate data-to-data hypotheses, [5] evaluates target-to-data hypotheses as in the PDA and JPDA algorithms. As partitions of the underlying probability space based upon the available information, the partition by target-to-data hypotheses is finer (sometimes considerably finer) than that by data-to-data hypotheses. This may cause both theoretical and practical problems. As discussed in detail in Part II of this report, the number of hypotheses poses always serious problems to any kind of multi-hypothesis tracking algorithms. One of the solutions is to combine or "aggregate" hypotheses. As mentioned in Sections 3 and 6.2, when hypotheses are combined, we may introduce cross-correlation among targets, which sometimes may require a large amount of memory. On the other hand, the infinite number of potential targets may cause some theoretical difficulty. For example, suppose a sensor observes a single measurement at one scan when no other data set has been generated and when we know there is no false alarm generated in this scan. Then we know at least one target has been born and detected. If we assume that the number of potential targets one of which may be the detected target is infinite, the probability of each target-to-data hypothesis is zero while we have one data-to-data hypothesis having probability one. Moreover, the assumption that each newly born target appears uniformly in a given space may be questionable.

7. CONCLUSIONS

A general multitarget tracking problem without *a priori* target identification (labeling) was formulated and a Bayesian solution to the problem was given. This formulation may be the most general within the assumption of no prior identification of targets. We did exclude, however, the possibilities of merged or split measurements. Inclusion of either one of these possibilities may cause considerable complication. We have shown that, when targets do not have *a priori* identification, we need to evaluate only data-to-data hypotheses and evaluation of target-to-track hypotheses is unnecessary. However, the consideration of target-to-track hypotheses was a key step in deriving our general algorithm. For the cases where several additional assumptions are made, called i.i.d. Poisson cases, the general result was reduced to a much simpler form. From this form, we can derive many specialized forms depending on the real problems.

We have also shown that, even when the algorithm is restricted to the i.i.d. Poisson cases, it includes most of the representative existing multitarget tracking algorithms as a proper subset in an appropriate sense, thereby providing a unified view on multitarget tracking problems. Besides generality, our theoretical and practical contributions include the following two aspects: First we considered the dependence of target detection on target states explicitly and incorporated it into our final form of the algorithm. This may be very important since, to the best knowledge of the authors, this dependence has been long neglected in the multitarget tracking literature while, in a practical sense, it is not negligible in the case of sensors with target state dependent detection characteristics. This will cover many practical issues such as radial velocity dependent detection by MTI-type radars, terrain masking, range and/or cross-section dependent detection by radars, radio-emission and/or target-activity dependent sensors, etc. Secondly we defined the likelihood of a measurement originating from a newly detected target in a very clear term. This provides the density of newly detected targets in Reid's algorithm with a transparent meaning. Moreover this eliminates a very ambiguous but often used notion of Poisson arrival of newly detected targets. Our basic viewpoint is that, whenever any target is newly detected by a sensor, it has been evading detection up to that point either because of failed detection or because of zero detection probability, i.e., because it has been out of field of view of any operating sensor, not active generating no detectable signal, not moving, thereby creating no radial velocity, etc.

As seen in Sections 4 and 5, the general multitarget tracking algorithm reduces the problem into a number of filtering problems that are subproblems, in either general or i.i.d.-Poisson cases. This means that all the filtering techniques developed up until now[43] are valuable sources for multitarget tracking. On the other hand, we may say that *hypothesis management* is one aspect truly unique to multitarget tracking. By hypothesis management[44], we mean many techniques to reduce the number of hypotheses while maintaining reasonable performance of the multitarget tracking algorithms. A variety of techniques have been developed in the past two decades. They provide us with very valuable information whenever implementational issues of any algorithm are concerned. For example, the "combining" of all the target-to-data hypotheses in PDA and JPDA algorithms can be viewed as a particular hypothesis management technique. At *A.I. & D.S.* , we have developed a general-purpose system for multitarget tracking, called GTC (generalized tracker and classifier)[45], which has been implemented in SAIL-, C- and LISP- programming languages. Implementational issues will be discussed in Part II of this report together with several examples created by GTC.

[43] See [22] for one of the most recent survey papers on this topic.

[44] Terminology borrowed from AI (artificial intelligence).

[45] The "tracker" part represents state estimation of geolocational entities while the "classifier" part any estimation of discrete entities. See [23] for a concise description of GTC.

REFERENCES

[1] Y. Bar-Shalom, "Tracking Methods in a Multitarget Environment," *IEEE Trans. Automat. Contr.*, Vol. AC-23, No. 4, pp.618-626, Aug. 1978

[2] H.L. Wiener, W.W. Willman, I.R. Goodman and J.H. Kullback, "Naval Ocean Surveillance Correlation Handbook, 1978," NRL Report 8340, Naval Research Lab., Washington D.C , 1979.

[3] H.L. Wiener, W.W. Willman, I.R. Goodman and J.H. Kullback, "Naval Ocean Surveillance Correlation Handbook, 1979," NRL Report 8402, Naval Research Lab., Washington D.C., 1980.

[4] D.B. Reid, "An Algorithm for Tracking Multiple Targets," *IEEE Trans. Automat. Contr.*, Vol. AC-24, No. 6, pp.843-854, Dec. 1979

[5] I.R. Goodman, "A General Model for the Contact Correlation Problem," NRL Report 8417, Naval Research Lab., Washington D.C., July, 1983.

[6] K.R. Pattipati, N.R. Sandell, Jr., and L.C. Kramer, "A Unified View of Multi-Object Tracking," Proc. 4th MIT/ONR Workshop on Distributed Information and Decision Systems Motivated by Command-Control-Communication Problems, Vol. 1, pp. 115-135, San Diego, CA., June 1981.

[7] G. Motheron, *Random Sets and Integral Geometry*, John Wiley & Sons., Inc., 1975.

[8] I.R. Goodman, " A General Model for the Multiple Target Correlation and Tracking Problem," Proc. 18th IEEE Conf. Decision and Contr., Fort Lauderdale, FL., pp. 383-388, Dec. 1979.

[9] R.P. Hughes and R.R. Tenney, "A Multiobject Tracking Algorithm for a Distributed Sensor Network," Proc. 3rd MIT/ONR Workshop on Distributed Information and Decision Systems, Sept. 1980.

[10] K.-C. Chang and Y. Bar-Shalom, "Joint Probabilistic Data Association for Multitarget Tracking with Possibly Unresolved Measurements," *IEEE Trans. Automat. Contr.*, Vol. AC-29, No. 7, pp. 585-594, July 1984.

[11] C.L. Morefield, "Application of 0-1 Integer Programming to Multitarget Tracking Problems," C, Vol. AC-22, No. 3, pp. 302-312, June 1977.

[12] D.L. Alspach and R.N. Lobbia, "A Score for Correct Data Association in Multi-Target Tracking," Proc. 18th IEEE Conf. Decision and Contr., Fort Lauderdale, FL., Dec. 1979.

[13] I.R. Goodman, "A Scoring Procedure for the Multiple Target Correlation and Tracking Problem," Proc. 19th IEEE Conf. Decision and Contr., Albuquerque, NM., pp. 829 - 834, Dec. 1980.

[14] D.B. Reid, "A Multiple Hypothesis Filter for Tracking Multiple Targets in a Cluttered Environment," Lockheed Report LMSC-D560254, Lockheed Palo Alto Research Laboratory, Palo Alto, CA., Sept., 1977.

[15] D.B. Reid and R.G. Bryson, "A Non-Gaussian Filter for Tracking Targets Moving over Terain," Conf. Record of 12th Asilomar Conf. on Circuits, Systems and Computers, pp. 112 - 117, Nov. 1978

[16] D.B. Reid, "The Application of Multiple Target Tracking Theory to Ocean Surveillance," Proc. of 18th IEEE Conf. Decision and Contr., Fort Lauderdale, FL., pp. 1046 - 1052, Dec. 1979.

[17] Y. Bar-Shalom and E. Tse, "Tracking in a Cluttered Environment with Probabilistic Data Association," *Automatica*, Vol. 11, pp. 451-460, Sept. 1975.

[18] Y. Bar-Shalom, T.E. Fortmann and M. Scheffé, "Joint Probabilistic Data Association for Multiple Targets in Clutter," Proc. the 1980 Conf. on Information Science and Systems, Princeton University, March 1980.

[19] T.E. Fortmann, Y. Bar-Shalom and M. Scheffé, "Multi-Target Tracking Using Joint Probabilistic Data Association," Proc. 19th IEEE Conf. Decision and Contr., Albuquerque, NM., pp. 807 - 812, Dec. 1980.

[20] J. Munkres, "Algorithm for the Assignment and Transportation Problems," *J. Soc. Indust. Appl. Math.*, Vol. 5, No. 1, pp. 32-38, March 1957.

[21] F. Bourgeois and J-C. Lassale, "An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices," *Communications of the ACM*, Vol. 14, No. 12, 1971.

[22] C.-B. Chang and J.A. Tabaczynski, "Application of State Estimation to Target Tracking," *IEEE Trans. Automat. Contr.*, Vol. AC-29, No. 2, pp. 98-109, Feb. 1984.

[23] Shozo Mori, Chee-Yee Chong, Edison Tse and Richard P. Wishner, "Generalized Tracker/Classifier (GTC) - A System for Tracking and Classification of Multiple Targets by Multiple Sensors," 6th MIT/ONR Workshop on Distributed Decision and Communication Problems, 1983.